B Flavour Tagging
with Artificial Neural Networks
for the CDF II Experiment

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Phys. Andreas Schmidt
aus Ludwigshafen am Rhein

Zusammenfassung

Motivation

Mesonen sind Teilchen, die aus einem Quark und einem Antiquark aufgebaut sind. Bestimmte neutrale Mesonen können sich in ihr Antiteilchen umwandeln, und umgekehrt. Die Rate, wie oft dieser Prozess in einer gegebenen Zeitspanne abläuft, wird *Oszillationsfrequenz* genannt.

Das Standardmodell der Teilchenphysik ist eine etablierte Theorie, in welcher die Eigenschaften der Elementarteilchen sowie deren Wechselwirkungen untereinander beschrieben werden. Wie jede Theorie muss sich auch das Standardmodell im Experiment bewähren. Eine Möglichkeit eines solchen Tests ist die Messung der Oszillationsfrequenz von B_s^0-Mesonen. Diese Messung bereitet auch den Weg für die Suche nach Physik jenseits des Standardmodells, die sich in einer Verletzung der CP-Symmetrie bei B_s^0-Meson-Zerfällen manifestieren kann.

Zur Zeit werden B_s^0-Mesonen nur an den beiden Experimenten CDF und DØ am Tevatron-Beschleuniger untersucht. Die Messung der B_s^0-Oszillationsfrequenz sowie die Suche nach CP-Verletzung im B_s^0-System sind zentrale Bestandteile des Flavour-Physik-Programms der CDF-Kollaboration.

Flavour Tagging

Die Messung der B_s^0-Oszillationsfrequenz besteht aus drei wesentlichen Schritten:

- Bestimmung des B_s^0-Flavours beim Zerfall des Mesons

- Messung der Flugstrecke und damit der Lebensdauer

- Bestimmung des Produktionsflavours des B_s^0-Mesons

Der dritte Punkt wird *Flavour Tagging* genannt. Als *Flavour* bezeichnet man im Allgemeinen die „Sorte" eines Quarks, also ob es sich zum Beispiel um ein s-Quark oder ein b-Quark handelt. Wie im vorliegenden Zusammenhang wird der Begriff auch zur Unterscheidung von Quarks und Antiquarks verwendet, was sich auch auf die aus ihnen
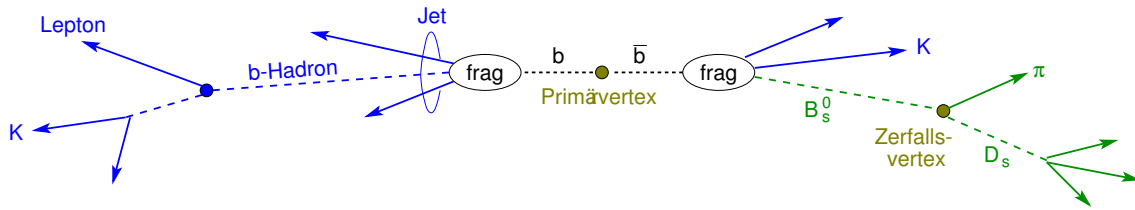
Abbildung 0.1: Schematische Darstellung eines typischen B_s^0 Zerfalls.

gebildeten Mesonen erstreckt. *Flavour Tagging* nennt man also ein Verfahren zur Unterscheidung zwischen B- und Anti-B-Mesonen. Das Ziel dieser Arbeit ist die Entwicklung eines verbesserten Flavour-Tagging-Algorithmus für das CDF-Experiment.

In Abbildung 0.1 ist der Zerfall eines B_s^0-Mesons schematisch dargestellt. Am Primärvertex, wo die Kollision der beschleunigten Teilchen stattfand, wird ein $b\bar{b}$-Quarkpaar erzeugt. Eines der beiden Quarks, in diesem Fall ein $\bar{b}$-Quark, bildet im Fragmentationsprozess ein B_s^0-Meson. Die entsprechende Seite des Ereignisses wird als *Same Side* bezeichnet. Das verbleibende Quark bildet auf der sogenannten *Opposite Side* ein weiteres B-Hadron.

Informationen über den Flavour des b-Quarks im B_s^0-Meson sind in verschiedenen Eigenschaften des Ereignisses enthalten, beispielsweise in den Zerfallsprodukten des Opposite Side B-Hadrons, in dessen Jet oder in Fragmentationsspuren auf der Same Side in der Nähe des B_s^0-Mesons.

Diese Informationen werden beim Flavour Tagging gesammelt und ausgewertet. Die bisher bei CDF verfügbaren Tagger konzentrieren sich in der Regel auf jeweils eine dieser Informationsquellen. Der hier vorgestellte Tagger vereint alle vorhandenen Hinweise in einem einzigen Werkzeug. Dabei werden künstliche Neuronale Netzwerke an vielen Stellen gewinnbringend eingesetzt.

Künstliche Neuronale Netzwerke

Künstliche Neuronale Netzwerke sind ein Klassifikationsverfahren für statistische Daten. Das Funktionsprinzip ist von der Natur abgeschaut und der Arbeitsweise unseres Gehirns nachempfunden.

In einem Lernprozess wird das Netzwerk mit Hilfe von Daten trainiert, bei denen die gewünschte Antwort vorab bekannt ist. Dabei werden die Struktur und die internen Parameter des Netzwerkes so eingestellt, dass die Eigenschaften des Trainingsdatensatzes möglichst gut abgebildet werden. Als Eingabegrößen dienen dabei Variablen, die einzeln oder in Kombination miteinander Unterscheidungskraft bezüglich der vorzunehmenden Klassifikation besitzen.

Die gelernten Eigenschaften werden in der Anwendung des Neuronalen Netzes benutzt, um die trainierte Entscheidung für Daten zu treffen, bei denen die Klassenzugehörigkeit

vorab nicht bekannt ist. Die Antwort wird dabei nicht als ja/nein-Entscheidung angegeben, sondern in Form einer Variablen, die als Wahrscheinlichkeit für die Sicherheit der Einordnung interpretiert werden kann. Ein künstliches Neuronales Netzwerk führt also eine Abbildung aus einem vieldimensionalen Eingaberaum auf eine eindimensionale Entscheidungsvariable durch.

Flavour–Tagging mit künstlichen Neuronalen Netzwerken

Fast alle bisher bei CDF verwendeten Flavour Tagger konzentrieren sich auf jeweils einen Teilaspekt, beispielsweise auf Leptonen oder auf die Ladung des Opposite Side Jets. Deswegen können sie nur in solchen Ereignissen eine Entscheidung liefern, in denen das spezifische Merkmal identifiziert werden kann, das sie zum Tagging benötigen. Im hier vorgestellten Ansatz werden verschiedene Informationsquellen zusammengeführt und gemeinsam ausgewertet. Eine Tagging-Entscheidung wird dabei für fast jedes Ereignis getroffen.

Das Verfahren besteht im Wesentlichen aus drei Stufen. Zuerst wird eine inklusive Rekonstruktion des Opposite-Side-B-Zerfallsvertex vorgenommen und weitere Größen vorbereitet, die für die folgenden Schritte von Nutzen sind. Danach wird für jede einzelne Spur der Informationsgehalt bestimmt, den sie über den Produktionsflavour des B_s^0-Mesons trägt. Zuletzt wird die Aussage aller Spuren in einem Ereignis zu einer Entscheidung auf Ereignisebene kombiniert.

Suche nach dem Sekundärvertex auf der Opposite Side

Die vollständige Rekonstruktion aller möglichen Zerfälle von B-Hadronen auf der Opposite Side ist vor allem aufgrund begrenzter Detektorakzeptanz nicht durchführbar. Dennoch wird versucht, den Zerfallsvertex räumlich einzugrenzen, um diese Informationen in späteren Schritten nutzen zu können.

Dazu wird ein erstes Neuronales Netzwerk darauf trainiert, Spuren zu finden, die mutmaßlich aus dem Zerfall des B-Hadrons stammen. Im Anschluss werden alle Spuren nach ihrer B-Herkunftswahrscheinlichkeit sortiert und aus den ersten vier paarweise Kandidaten für den gesuchten Zerfallsvertex generiert. In einem zweiten Neuronalen Netzwerk werden die Vertexkandidaten bewertet. Der Vertex mit der höchsten Wahrscheinlichkeit, der wahre Zerfallsvertex zu sein, wird in den nächsten Schritten weiterverwendet. Mit Hilfe des nun bekannten Sekundärvertex wird in einem dritten Neuronalen Netzwerk die Entscheidung, welche Spuren aus dem B-Zerfall stammen, wiederholt und verfeinert.

Weiterhin werden in diesem Schritt einzelne Spuren zu *Jets* gruppiert und Verfahren zur Identifikation von Elektronen und Myonen angewandt. Mit einem Neuronalen Netzwerk werden Leptonen, die wahrscheinlich aus dem Opposite-Side-B-Zerfall stammen, von übrigen im Detektor vorhandenen Leptonen unterschieden.
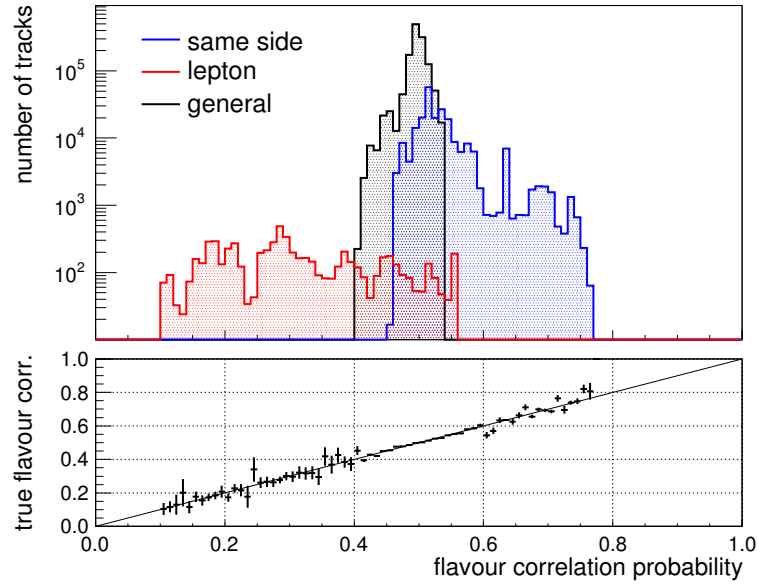
Abbildung 0.2: Ergebnis des spurbasierten Flavour Tagging in den drei Spurklassen. Die Darstellung wurde logarithmisch gewählt, da die meisten Spuren nur kleine Beträge liefern und sich daher in der Mitte der Verteilung ansammeln.

Spurbasiertes Flavour Tagging

Im zweiten Schritt wird für jede Spur die Korrelation zwischen ihrer Ladung und dem Produktionszustand des Same-Side-B_s^0-Mesons mit Neuronalen Netzwerken bestimmt. Weil Flavour-Information aus verschiedenartigen Quellen gewonnen werden kann und jeder dieser Effekte für sich genommen klein ist, werden die Spuren in drei Klassen aufgeteilt, um mit spezialisierten Netzwerken alle Informationsquellen optimal zu nutzen.

Die erste Klasse wird von den Spuren gebildet, die zum selben Jet gehören wie die Spuren, aus denen das Same-Side-B_s^0-Meson rekonstruiert wurden. In dieser *Same-Side-Spurklasse* erwartet man überwiegend positive Korrelationen zwischen Spurladung und Flavour. Zur *Lepton-Spurklasse* gehören alle Spuren, die nicht in der Same-Side-Spurklasse enthalten sind und vom B-Lepton-Identifikationsnetz aus dem ersten Schritt als gute Lepton-Kandidaten eingestuft wurden. Für diese Klasse erwartet man vorwiegend Antikorrelationen zwischen Ladung und Same-Side-B_s^0-Flavour. Alle übrigen Spuren bilden die *allgemeine Spurklasse*, in der ohne gesonderte Vorauswahl alle verbleibenden Effekte ausgewertet werden. Da für die deutlichsten Flavour-Informationsquellen bereits in den beiden ersten Klassen gesondert behandelt wurden, ist in dieser Klasse nicht mit hohen Flavour-Korrelationen zu rechnen.

Das Ergebnis dieser Einordnung ist in Abbildung 0.2 gezeigt. Die meisten Spuren tragen kaum Information, während anhand einiger Spuren eine sehr sichere Aussage möglich ist.
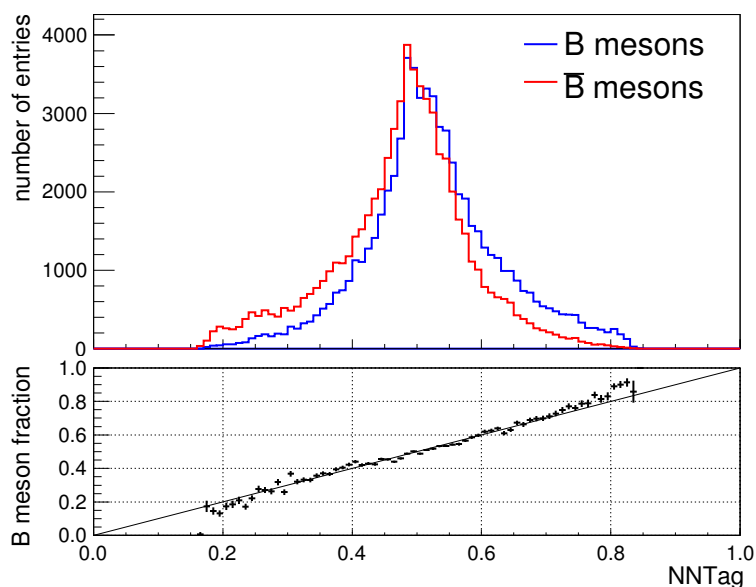
Abbildung 0.3: Ausgabe des Neuronalen Netzwerks zur Bestimmung des B_s^0-Produktionsflavours, separat für wahre B- und $\bar{B}$-Mesonen eines simulierten Datensatzes. In der unteren Darstellung ist der wahre B-Meson-Anteil als Funktion der Netzwerkvorhersage dargestellt. Das Verhalten kommt dem durch die diagonale Linie dargestellten Idealfall eines optimalen Netzwerks sehr nah.

Kombination auf Ereignisebene

In jeder Spurklasse werden alle Spuren durch Multiplikation ihrer Korrelationswahrscheinlichkeiten kombiniert. Die entstandene Größe wird danach so normiert, dass sie eine Wahrscheinlichkeitsaussage für den Flavour auf Ereignisebene darstellt. Zusätzlich werden die Spuren aller drei Klassen in gleicher Weise zu einer Aussage über den B_s^0-Flavour kombiniert.

Die so gewonnenen Flavour-Aussagen werden mit einem Neuronalen Netzwerk zusammengefasst. Als weitere Eingabegröße für dieses Netzwerk wird ein Schätzwert für die Ladung des Opposite Side B-Hadrons verwendet. Die Ausgabe dieses Netzwerkes ist in Abbildung 0.3 zu sehen. Im zentralen Bereich sammeln sich Ereignisse, bei denen eine sichere Bestimmung des Flavours nicht möglich ist. Die Verteilungen für beide Klassen zeigen dennoch einen deutlichen Unterschied, und für einige Ereignisse wird sogar eine Zuverlässigkeit von mehr als 80% erreicht.

Validierung

Die Leistungsfähigkeit eines Flavour Taggers wird durch die Größen *Dilution D* und *Effizienz ϵ* beschrieben. Die Dilution ist ein Maß für den Anteil richtiger Entscheidungen. Die Effizienz misst den Anteil der Ereignisse, für die überhaupt eine Entscheidung getroffen wurde. Als Tagging-Leistung $\mathcal{T}$ wird üblicherweise $\mathcal{T} = \epsilon D^2$ angegeben, weil dies der effektiven Größe des mit Tagging-Entscheidungen versehenen Datensatzes entspricht und die Signifikanz der B_s^0-Oszillationsmessung von dieser Kombination abhängt.

Der vorliegende Flavour-Tagger liefert eine Entscheidung für jedes Ereignis, in dem mindestens eine Spur zusätzlich zum rekonstruierten B_s^0-Meson enthalten ist und erreicht damit eine Effizienz von mehr als 99%. Seine Tagging-Leistung auf dem simulierten Datensatz beträgt

$$\mathcal{T} = 4.6\%. \tag{0.1}$$

Die Entwicklung des Flavour-Taggers erfolgte auf simulierten B_s^0-Ereignissen, da hier im Gegensatz zu echten Daten die für Trainingsprozesse nötige Wahrheitsinformation zugänglich ist. Aus physikalischen und technischen Gründen ist die Realitätstreue solcher Simulationen nicht perfekt. Das Verhalten des damit entwickelten Taggers bei Anwendung auf echte Daten muss daher überprüft werden. Insbesondere ist dabei von Interesse, ob der Tagger die Sicherheit einer Entscheidung korrekt vorhersagt und somit sein Ausgabewert als Wahrscheinlichkeit interpretiert werden kann.

Ein Vergleich zwischen Eigenschaften der Simulation und der Daten zeigte Unterschiede in der Beschreibung des Triggers. Die Datenrate beim Auslesen des CDF-Detektors ist begrenzt, und Triggerszenarien für verschiedene Analysebedürfnisse konkurrieren hier. Um die verfügbare Bandbreite optimal zu nutzen, wird die Vergabe an einzelne Triggerszenarien während der Datennahme dynamisch optimiert. Da dies ein hochkomplexer Prozess ist, der von vielen Betriebsparametern abhängt und sich zeitlich entwickelt, ist er sehr schwer zu modellieren und wird nicht simuliert. Im simulierten Datensatz lag daher eine andere Verteilung auf die drei verwendeten Triggerszenarien vor als im Datensatz. Diese Diskrepanz wurde mit Gewichten korrigiert.

Geladene B^+- sowie neutrale B^0-Mesonen bieten weitere Möglichkeiten zur Validierung des Flavour-Taggers. Aufgrund physikalischer Unterschiede bei den möglichen Quellen von Flavour-Information ist nicht zu erwarten, dass ein B_s^0-Tagger für andere B-Mesonen optimale Entscheidungen liefert. Simuliert man jedoch einen Datensatz mit dem gewünschten Meson-Typ und wiederholt das Training der betroffenen Neuronalen Netzwerke im Tagger, kann man auf diese Weise sowohl das Simulationsverfahren als auch die Tagging-Prozedur überprüfen.

Die Erhaltung elektrischer Ladung verbietet es B^+-Mesonen, mit ihren Antiteilchen zu oszillieren. Daher ist ihr Produktions- und Zerfallsflavour identisch. Während es also aus Analysesicht keinen Bedarf für Flavour-Tagging gibt, bieten sich für dessen Validierung auf echten Daten gute Voraussetzungen, da die gewünschte Wahrheitsinformation auch

bei Daten bekannt ist. Die Validierungsergebnisse für die B^+-Variante des vorliegenden Flavour-Taggers zeigen sehr gute Übereinstimmung zwischen vorhergesagter und gemessener Dilution. Damit bestehen keine Hinweise auf größere Unstimmigkeiten zwischen Simulation und Daten. Das Ergebnis stützt auch die Funktionsfähigkeit der verwendeten Tagging-Strategie.

Die Oszillationen von B^0-Mesonen sind denen der B_s^0-Mesonen physikalisch sehr ähnlich. Aufgrund der niedrigeren Oszillationsfrequenz und der Größe des verfügbaren Datensatzes sind sie jedoch experimentell besser handhabbar. Mit der B^0-Variante des Taggers konnte die B^0-Oszillationsfrequenz in Übereinstimmung mit dem Weltmittelwert gemessen werden. Die beobachtete Amplitude

$$\mathcal{A} = 0.92 \pm 0.08 \tag{0.2}$$

ist innerhalb des statistischen Fehlers verträglich mit der physikalisch richtigen Oszillationsamplitude $\mathcal{A}_{\mathrm{phys}} = 1$. Damit bestätigt auch die B^0-Validierung die Richtigkeit der vorhergesagten Dilution.

Die einzige Möglichkeit zur Validierung des B_s^0-Taggers mit echten Daten ist die Messung der B_s^0-Oszillationsfrequenz. Während die volle Mischungsanalyse mit B_s^0-Mesonen aus mehreren Zerfallskanälen noch andauert, konnten erste vorläufige Ergebnisse mit dem wichtigsten Kanal zur Validierung dieses Taggers herangezogen werden. Auch in diesem Fall konnte die Oszillationsfrequenz in Übereinstimmung mit dem Weltmittelwert gemessen werden. Die gemessene Amplitude

$$\mathcal{A} = 0.98 \pm 0.28 \tag{0.3}$$

liegt ebenfalls sehr nahe beim erwarteten Wert 1. Der statistische Fehler dieser Messung wird mit Abschluss der Analyse aller Zerfallskanäle verbessert werden.

Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurde ein neuer B_s^0-Flavour-Tagger für das CDF-Experiment entwickelt und sowohl mit Hilfe von B^+- und B^0-Mesonen als auch in einer vorläufigen Analyse von B_s^0-Oszillationen mit dem wichtigsten B_s^0-Zerfallskanal validiert. Im Gegensatz zu bisher vorhandenen Taggern werden alle Flavour-Informationsquellen in einem einheitlichen Verfahren ausgewertet und kombiniert. Dabei wird an vielen Stellen Gebrauch von künstlichen Neuronalen Netzwerken gemacht.

Die vorliegende Version des Tagger macht keinen Gebrauch von Informationen zur Teilchenidentifikation, die beim CDF-Experiment durch Flugzeitmessung oder spezifischen Energieverlust im Spurdetektor gewonnen wird. Während die damit erreichbare Trennung von Kaonen und Pionen besonders beim B_s^0-Same-Side-Tagging von Nutzen wäre,

gab und gibt es immer wieder technische und organisatorische Schwierigkeiten mit der Verfügbarkeit der entsprechenden Größen und deren Simulation. Üblicherweise vergeht eine bedeutende Zeitspanne zwischen Datennahme und Kalibration der Teilchenidentifikation.

In einer Studie wurde ermittelt, wie groß das Verbesserungspotential für die Tagging-Leistung durch Teilchenidentifikation ist. Die entsprechende Version des Taggers erreicht eine Leistung von

$$\mathcal{T} = 6.7\% \tag{0.4}$$

auf dem simulierten Datensatz, was eine Verbesserung von gut 40% gegenüber der jetzigen Version in Aussicht stellt.

Bisher wurden bei CDF ein kombinierter Opposite-Side-Tagger mit einer Leistung $\mathcal{T} = 1.8\%$ und ein Same-Side-Kaon-Tagger mit $\mathcal{T} = 3.7\%$ verwendet, wobei letzterer auf Teilchenidentifikation angewiesen ist. Die kombinierte Leistung beider Tagger ist nicht angegeben, kann aber nicht höher sein als die Summe der beiden.

Der hier vorgestellte neue Tagger bietet also eine deutliche Erhöhung der Tagging-Leistung für Daten, für die noch keine Kalibration der Teilchenidentifikation durchgeführt wurde. Ist diese zusammen mit einer verlässlichen Simulation verfügbar, kann der neue Tagger mit geringem Aufwand eine Leistung erreichen, die voraussichtlich 20% über der bisher verfügbaren kombinierten Leistung beider Tagger liegt.

Die finale Validierung des Verfahrens anhand der vollständigen B_s^0-Oszillationsanalyse mit allen verfügbaren Daten und Zerfallskanälen wird zugleich die erste physikalische Anwendung des neuen Taggers sein. Danach kann er in der Suche nach CP-Verletzung in $B_s^0 \to J/\psi\,\phi$-Zerfällen eingesetzt werden, die eine der wichtigsten Analysen im Flavour-Physik-Programm des CDF-Experimentes ist. Auch für weitere Suchen nach CP-Verletzung in $B_s^0 \to J/\psi\,f_0(980)$ und in $B_s^0 \to K^+ K^-$-Zerfällen ist Flavour-Tagging ein zentraler Bestandteil.

Contents

List of Figures

1 Introduction

One of the central questions arising from human curiosity has always been what *matter* is ultimately made of, with the idea of some kind of elementary building-block dating back to the ancient greek philosophers. Scientific activities of multiple generations have contributed to the current best knowledge about this question, the *Standard Model* of particle physics. According to it, the world around us is composed of a small number of stable elementary particles: Electrons and two different kinds of quarks, called *up* and *down* quarks. Quarks are never observed as free particles, but only as bound states of a quark-antiquark pair (*mesons*) or of three quarks (*baryons*), summarized as *hadrons*. Protons and Neutrons, the constituents forming the nuclei of all chemical elements, are baryons made of up and down quarks. The electron and the electron *neutrino* – a nearly massless particle without electric charge – belong to a group called *leptons*. These two quarks and two leptons represent the first *generation* of elementary particles.

There are two other generations of particles, which seem to have similar properties as the first generation except for higher masses, so there are six quarks and six leptons altogether. They were around in large amounts shortly after the beginning of the universe, but today they are only produced in high energetic particle collisions. Properties of particles are described by *quantum numbers*, for example charge or spin. For every type of particle, a corresponding *antiparticle* exists with the sign of all charges swapped, but similar properties otherwise.

The Standard Model is a very successful theory, describing the properties of all known particles and the interactions between them. Many of its aspects have been tested in various experiments at very high precision. Although none of these experimental tests has shown a significant deviation from the corresponding Standard Model prediction, the theory can not be complete yet: Cosmological aspects like gravity, dark matter and dark energy are not described, and open questions remain in the sector of neutrino masses and neutrino oscillations. Also no answer has been given to the question of matter-antimatter asymmetry observed in the contemporary universe. Assuming that the Big Bang created equal amounts of matter and antimatter, there must be effects where nature treats matter and antimatter somehow different. This can happen through a mechanism called *CP violation*, which has been observed within the Standard Model, but not in the necessary order of magnitude. For all these reasons, the search for *New Physics* – theories beyond the Standard Model – is one of the main objectives of modern particle physics. In this global effort, *flavour physics* is the field of transitions between the different types of quarks, called *quark flavours*, wherein the examination of *B meson oscillations* and the search for

CP violation in B_s^0 meson decays set the stage for the work presented in this thesis.

Neutral B mesons can transform themselves to $\bar{B}$ mesons and vice versa. This process is called *oscillation* and happens with a distinct *oscillation frequency*. Depending on the individual time between production and decay, a B meson can decay as a $\bar{B}$ meson with a certain probability. The value of the oscillation frequency is governed by parameters of the quark level transformation process. Measuring the frequency provides an experimental test of its theoretical understanding. The oscillation frequency for B^0 mesons has been precisely measured by several experiments, with most important contributions made by the B factory experiments BaBar [1, 2] and Belle [3, 4]. Until the Large Hadron Collider at CERN takes up full operations, the *Tevatron* – the world's second most powerful particle accelerator – is the only place where large samples of B_s^0 mesons are produced and studied. The B_s^0 oscillation frequency has been measured by its two experiments CDF [5] and DØ [6] about three years ago. Since then, the available amount of data has been increased roughly by a factor of five, opening up the possibility for an updated measurement at increased statistical precision.

Measuring the oscillation frequency needs three main ingredients: The flavour of the B meson at decay time, the time between production and decay as well as the flavour at production time. The first two aspects are experimentally well under control and mostly limited by detector performance, while the main challenge is in the third point. The determination of the production flavour – whether a reconstructed B meson was a B or $\bar{B}$ at production time – is called *flavour tagging*, and the development of an advanced flavour tagging algorithm for the CDF experiment is the objective of this work. Indirect information about the production flavour can be gained from various sources, and former CDF flavour tagging algorithms were focused on one source each. The new flavour tagger presented here is designed to evaluate all available information in a uniform way, take advantage of correlations between the individual hints and combine all knowledge into a single flavour tag. The good classification performance of artificial neural networks is extensively used in the process.

The new flavour tagger can be used for an improved measurement of the B_s^0 oscillation frequency, which is at the same time the only possibility to validate the tagging performance on data. After successful validation, the main application of improved flavour tagging are measurements of time dependent CP violation in B_s^0 meson decays. Besides the current cornerstone of the current CDF flavour physics program, the analysis of CP violation in $B_s^0 \rightarrow J/\psi\,\phi$ decays [7], also future searches for New Physics in $B_s^0 \rightarrow J/\psi\,f_0(980)$ or $B_s^0 \rightarrow K^+K^-$ decays depend on flavour tagging.

This thesis is organized as follows: In chapter two, a brief overview on the theory of B_s^0 oscillations and the analysis to measure the oscillation frequency is given and the physics of flavour tagging is explained. In the third chapter, the Tevatron accelerator at Fermilab and the CDF detector are described. The fourth chapter makes an introduction to classification of statistical data and the principle of artificial neural networks. Chapter

five contains a detailed description of the new combined flavour tagging algorithm. Studies for validating the new tagger on simulated events as well as on data are presented in chapter six. The last chapter gives a summary of development and validation as well as an outlook to application and possible further improvements of the new flavour tagger.

2 Theoretical Overview

2.1 Flavour Physics and B_s^0 Oscillations

2.1.1 The CKM Matrix

Flavour physics is the field of transitions between different types of quarks, called *quark flavours*. In the Standard Model of particle physics, quark mass eigenstates and flavour eigenstates are not the same. The transformation

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\mathrm{CKM}} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.1}$$

between the mass (d, s, b) and flavour (d', s', b') base is represented by a complex 3×3 matrix V_{CKM}, the *Cabibbo-Kobayashi-Maskawa* (CKM) matrix. It was introduced in 1973 by Kobayashi and Maskawa [8], based on the transformation for four quarks by Cabibbo [9], and predicted the existence of three generations of matter at a time when only three quarks were known experimentally. Half of the Nobel Prize in Physics 2008 was awarded to Kobayashi and Maskawa *for the discovery of the origin of the broken symmetry which predicts the existence of at least three families of quarks in nature* [10]. In case of not more than three generations, the CKM matrix is unitary and fulfils the relation

$$V_{\mathrm{CKM}}^\dagger V_{\mathrm{CKM}} = \mathbb{1} = V_{\mathrm{CKM}} V_{\mathrm{CKM}}^\dagger, \tag{2.2}$$
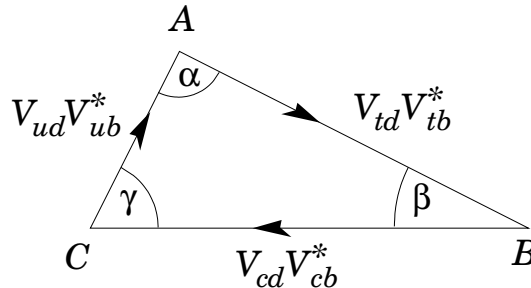
Figure 2.1: The unitarity triangle derived from the unitarity condition of the CKM matrix, rotated that $V_{cd}V_{cb}^*$ is on the real axis.
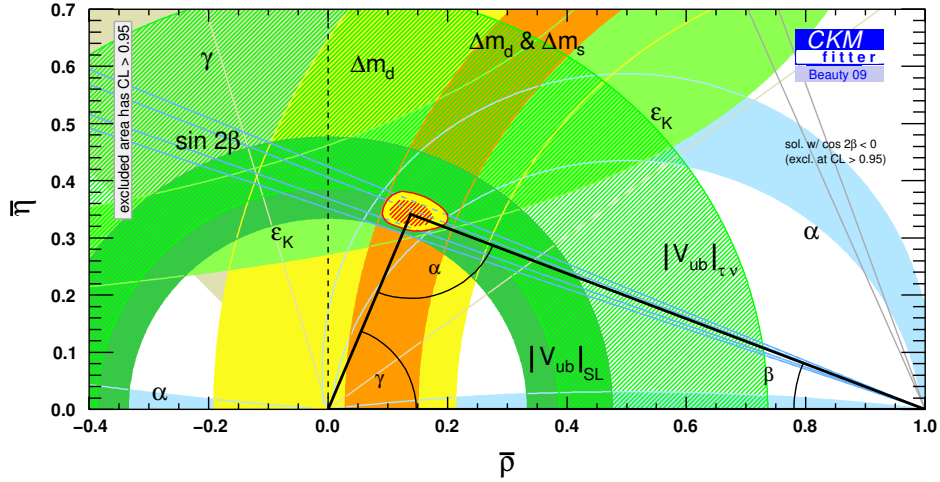
Figure 2.2: Current status in the determination of CKM triangle parameters [12].

which implies constraints between the matrix elements and therefore reduces the matrix to three real parameters and one imaginary phase. A popular parametrization was introduced by Wolfenstein [11], using the four real parameters λ, A, ρ and η:

$$V_{\text{CKM}} \approx \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & \lambda^3 A(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & \lambda^2 A \\ \lambda^3 A(1 - \rho - i\eta) & -\lambda^2 A & 1 \end{pmatrix} + O(\lambda^4). \tag{2.3}$$

The unitarity condition from equation 2.2 can be visualized in form of triangles in the complex plane. Applying the condition to the first and third column of V_{CKM} yields the equation

$$V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{tb}^* V_{td} = 0, \tag{2.4}$$

which is the most common representation of a *unitarity triangle*. Figure 2.1 shows a graphical representation of the triangle, rotated in a way that $V_{cd}V_{cb}^*$ coincides with the real axis. The side lengths and the three angles of the triangle can be determined from various measurements. By overconstraining the triangle parameters, the consistency of all involved measurements and the underlying theory is tested. Any significant disagreement would result in an unclosed triangle and be an indicator for New Physics. The current status of this effort is summarized in figure 2.2.

2.1.2 B_s^0 Meson Oscillations

Transitions between particles and their antiparticles, called *mixing*, have been observed for neutral kaons in the 1960s. In the B^0 system, mixing was observed in 1987 [13], and

the mass difference Δm_d has since been measured precisely [1, 2, 3, 4].

The time evolution of the B_s^0 flavour eigenstates

$$|B_s^0\rangle = |\bar{b}s\rangle, \; |\bar{B}_s^0\rangle = |b\bar{s}\rangle \tag{2.5}$$

is given by the Schrödinger equation

$$i\frac{\partial}{\partial t}\left(\begin{array}{c} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{array} \right) = \left(\mathbf{M} - \frac{i}{2}\mathbf{\Gamma}\right)\left(\begin{array}{c} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{array} \right) \tag{2.6}$$

with the hermitian 2×2 mass and decay matrices $\mathbf{M}$ and $\mathbf{\Gamma}$. Diagonalization of this effective Hamiltonian leads to the heavy and light *mass eigenstates*:

$$|B_s^H\rangle = p\,|B_s^0\rangle - q\,|\bar{B}_s^0\rangle, \tag{2.7}$$
$$|B_s^L\rangle = p\,|B_s^0\rangle + q\,|\bar{B}_s^0\rangle. \tag{2.8}$$

The complex parameters p and q satisfy the relation $|p|^2 + |q|^2 = 1$. Mass and width difference between the heavy and light mass eigenstates are defined as

$$\Delta m = m_H - m_L = 2\,|M_{12}|, \tag{2.9}$$
$$\Delta\Gamma = \Gamma_L - \Gamma_H = 2\,|\Gamma_{12}|\cos(\phi_s), \tag{2.10}$$

where ϕ_s is a complex phase between M_{12} and Γ_{12}:

$$\phi_s = \arg\left(-\frac{M_{12}}{\Gamma_{12}}\right). \tag{2.11}$$

Due to the mass difference, the heavy and light eigenstates have different time evolutions

$$|B_s^H(t)\rangle = |B_s^H(0)\rangle\,e^{-im_H t}e^{-t/\tau}, \tag{2.12}$$
$$|B_s^L(t)\rangle = |B_s^L(0)\rangle\,e^{-im_L t}e^{-t/\tau}, \tag{2.13}$$

where a difference in lifetime is neglected for simplicity of the equations. The difference in time evolution results in a time-dependent B_s^0 - $\bar{B}_s^0$ oscillation with a frequency equal to the mass difference Δm_s. Therefore a $|B_s^0\rangle$ state can decay either as $|B_s^0\rangle$ or $|\bar{B}_s^0\rangle$ depending on the time of its decay, following the probability densities

$$P(B_s^0 \rightarrow B_s^0,\, t) = \frac{\Gamma_s}{2}\,e^{-\Gamma_s t}(1 + \cos(\Delta m_s t)), \tag{2.14}$$

$$P(B_s^0 \rightarrow \bar{B}_s^0,\, t) = \frac{\Gamma_s}{2}\,e^{-\Gamma_s t}(1 - \cos(\Delta m_s t)), \tag{2.15}$$

where $\Gamma_s = 1/\tau_s$ denotes the mean decay width and τ_s the mean lifetime of the two mass
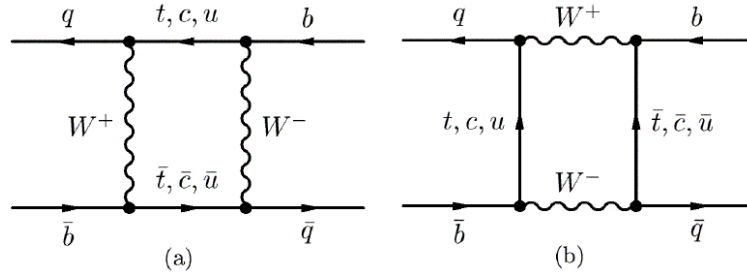
Figure 2.3: The most important Feynman graphs for $B_q^0 \to \overline{B}_q^0$ transitions, q denoting a d or a s quark. In the B_s^0 case, the diagrams with the charm or up quarks are suppressed [14].

eigenstates.

The two most important Feynman diagrams for neutral B meson mixing are shown in figure 2.3 and can be used to calculate the oscillation frequency [15]. Besides various parameters for handling electroweak and QCD effects, the oscillation frequency is proportional to $|V_{tb}^* V_{td}|^2$ in the B^0 and to $|V_{tb}^* V_{ts}|^2$ in the B_s^0 case. If mass and flavour eigenstates were identical, the off diagonal elements V_{td} and V_{ts} would be zero and no mixing would occur. Expressing the CKM Matrix elements in the Wolfenstein parametrization (eq. 2.3) and assuming no other significant differences between B^0 and B_s^0, the ratio of the two frequencies can be expressed as

$$\frac{\Delta m_s}{\Delta m_d} \approx \frac{|V_{ts}|^2}{|V_{td}|^2} \approx \frac{1}{\lambda^2} \approx 20. \qquad (2.16)$$

This means that B_s^0 oscillations are significantly faster than B^0 oscillations, which leads to additional challenges for their experimental analysis.

2.1.3 CP Violation in the B_s^0 Meson System

The CP transformation is the combination of charge conjugation C, exchanging particles and antiparticles, and parity transformation P, reversing the handedness of space. The laws of nature are identical for matter and antimatter if CP is an exact symmetry. Gravity, the electromagnetic and the strong interaction preserve the C and P symmetries individually and are therefore as well CP-symmetric. The weak interaction on the other hand violates both C and P symmetries in the maximum possible way by only coupling to left-handed fermions and right-handed anti-fermions, while still preserving CP in most processes.

Violation of the CP symmetry has been discovered 1964 in neutral Kaon decays [16] and also been established in the B meson system. In the Standard Model, the violation of the CP symmetry is caused by complex phases in the Yukawa coupling of quarks to

the scalar Higgs field, which can be reduced to one complex phase in the CKM matrix by eliminating unphysical phases [8]. In the Wolfenstein parametrization (eq. 2.3), this phase corresponds to the parameter η.

In the B_s^0 system, the search for CP violation in $B_s^0 \to J/\psi\,\phi$ decays is a sensitive probe of New Physics and therefore one of the most active analysis fields in the current CDF B physics program [7]. In a time dependent measurement of the CP asymmetry

$$A_{CP}(t) = \frac{\Gamma(\bar{B}_s^0 \to f_{CP}) - \Gamma(B_s^0 \to f_{CP})}{\Gamma(\bar{B}_s^0 \to f_{CP}) + \Gamma(B_s^0 \to f_{CP})} \approx \pm sin(2\beta_s)\ sin(\Delta m_s\ t), \qquad (2.17)$$

it gives access to the angle $\beta_s^{J/\psi\,\phi}$ of the unitarity triangle given by the second and third column of the CKM matrix:

$$2\ \beta_s^{J/\psi\,\phi} = -\arg[(V_{tb}V_{ts}^*)^2/(V_{cb}V_{cs}^*)^2] \qquad (2.18)$$

The presence of New Physics enhancing ϕ_s (eq. 2.11) would affect $\beta_s^{J/\psi\,\phi}$ in the same way [17]. As the Standard Model prediction $\phi_s = 0.004$ is very small [18], the measurement of a significantly larger value would be a clear indication of New Physics.

The $J/\psi\,\phi$ final state can be reached with and without mixing and is a mixture of CP even and odd states. In the decay of the pseudoscalar B_s^0 meson into two vector mesons J/ψ and ϕ, the $L = 0, 2$ states are CP even and the $L = 1$ state is CP odd. The sensitivity to the phase is increased by studying the time evolution of CP even and CP odd states separately, which can be distinguished by the angular distributions of their decay products. Information about mixing is gained by tagging the production flavour of the B_s^0 meson.

2.2 Outline of the B_s^0 Mixing Analysis

From an experimentalist's point of view, the analysis of B_s^0 oscillations is done by measuring the time dependent asymmetry

$$A(t) = \frac{N_{B_s^0 \to B_s^0}(t) - N_{B_s^0 \to \bar{B}_s^0}(t)}{N_{B_s^0 \to B_s^0}(t) + N_{B_s^0 \to \bar{B}_s^0}(t)} \qquad (2.19)$$

between the number of B_s^0 mesons decaying as B_s^0 mesons ($N_{B_s^0 \to B_s^0}$) and those decaying as $\bar{B}_s^0$ mesons ($N_{B_s^0 \to \bar{B}_s^0}$). Also the charge conjugate situation where a $\bar{B}_s^0$ was produced is implied, but not included in the equations for simplicity. The time dependence of the

single decay rates is

$$N_{B_s^0 \to B_s^0}(t) = N_0 \cdot \frac{e^{-t/\tau}}{2\tau}(1 + \cos \Delta m_s t), \qquad (2.20)$$

$$N_{B_s^0 \to \bar{B}_s^0}(t) = N_0 \cdot \frac{e^{-t/\tau}}{2\tau}(1 - \cos \Delta m_s t), \qquad (2.21)$$

where N_0 is the number of B_s^0 mesons at $t = 0$ with t being the time in the B_s^0 rest frame. Combining these equations leads the asymmetry $A(t)$ as a function of Δm_s

$$A(t) = \cos(\Delta m_s t). \qquad (2.22)$$

Making a time dependent measurement of this asymmetry requires three main ingredients:

- flavour at decay time of the B_s^0 meson

- decay time of the B_s^0 meson in its rest frame

- flavour at production time of the B_s^0 meson

The first two items are covered by the reconstruction of a B_s^0 candidate, the third step is called *flavour tagging* and is the main objective of this work. An overview of approaches to flavour tagging is given later in this chapter. The significance of the B_s^0 mixing signal depends on three components [19]:

$$\text{signi} = \sqrt{\frac{\epsilon D^2}{2}} \, \exp\left(-\frac{(\Delta m_s \cdot \sigma_{c\tau})^2}{2}\right) \, \frac{S}{\sqrt{S + B}} \qquad (2.23)$$

The three terms correspond to important parts of the mixing analysis: $S/\sqrt{S + B}$ is optimized during B_s^0 signal reconstruction and selection, $\sigma_{c\tau}$ depends on the decay time reconstruction, and ϵD^2 is the contribution from flavour tagging. These components are described in the next paragraphs.

2.2.1 Signal Reconstruction

The reconstruction of a B_s^0 meson is called *exclusive* if all particles originating from the decay are found. If a particle is missing, only a *partial* reconstruction is possible. At CDF, the exclusive reconstruction of a B_s^0 meson is possible if all final state particles are charged hadrons. B_s^0 mesons in the *semileptonic* decay channel $B_s^0 \to l^+ \nu_l X$ can only be reconstructed partially because neutrinos are not visible in the detector. Due to the absence of a track, the neutrino momentum is unknown, so momentum resolution for this channel is significantly worse than for the hadronic decays.

The most important hadronic B_s^0 decay channel is $B_s \to D_s \pi$. For development of the tagger presented in this work on simulated events and for the first validation on data, the decay

$$B_s^0 \to D_s^\pm \pi^\mp, \; D_s^\pm \to \phi \pi^\pm, \; \phi \to K^+ K^- \tag{2.24}$$

is used. Of all fully reconstructed flavour specific B_s^0 decay channels available at CDF, it has the largest signal yield and is therefore also referred to as the *golden channel*.

The flavour of the B_s^0 meson at decay time is determined from the charge of its decay products, for example the pion charge from the $B_s^0 \to D_s^\pm \pi^\mp$.

2.2.2 B_s^0 Decay Time Measurement

In comparison to most other unstable particles, B mesons are relatively long lived because they decay only via the weak interaction, and the V_{cb} element of the CKM matrix is small. The lifetime of a B_s^0 meson is about 1.5 ps [14]. At CDF, a B_s^0 meson typically travels several hundred micrometers before decay, which is too short to reach the detector or even leave the beam pipe, but long enough to find the displaced B decay vertex.

The decay time of the B meson in its rest frame, its *proper decay time* t, is given by

$$t = \frac{L}{\beta \gamma \; c} = L \; \frac{m_B}{c \; p(B)} \tag{2.25}$$

where L is the flight distance, $\beta \gamma = p(B)/m_B$ the relativistic boost, $p(B)$ the momentum of the B meson and m_B its mass. As in many experiments, the detector resolution in the $x - y$ plane is better than along the z axis, therefore often the transverse projection

$$t = L_{xy} \; \frac{m_B}{c \; p_T(B)} \tag{2.26}$$

is used. In this expression, L_{xy} and p_T are the $x-y$ projections of the respective quantities.

The error σ_t on the proper time t has contributions from the resolutions $\sigma_{L_{xy}}$ of the decay length and σ_{p_T} of the transverse momentum:

$$\sigma_t = \sqrt{\left(\frac{m_B}{c \; p_T}\right)^2 \cdot \sigma_{L_{xy}}^2 + \left(\frac{t}{p_t}\right)^2 \cdot \sigma_{p_T}^2} \tag{2.27}$$

The contribution of σ_{p_T} can be neglected in exclusively reconstructed hadronic decays, but is an important term in semileptonic decays.
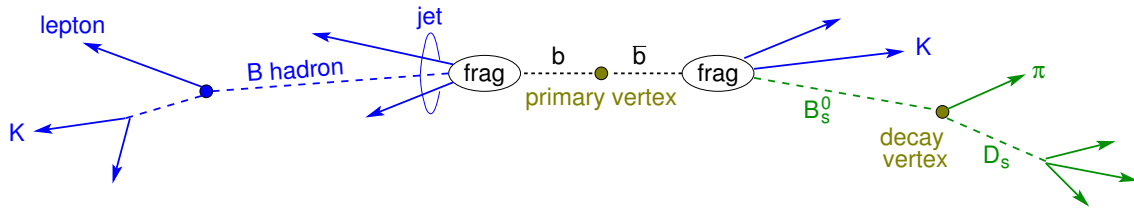
Figure 2.4: Illustration of a typical B_s^0 event.

2.2.3 Amplitude Scan Method

An alternative to directly fitting for the asymmetry $A(t)$ is the *amplitude scan method* [19]. Even if no mixing signal is seen, it is good for setting limits on a mixing frequency. A new parameter $\mathcal{A}$, called *amplitude*, is introduced:

$$A(t) \sim \mathcal{A} \cdot D \cdot \cos(\Delta m_s t). \tag{2.28}$$

Several asymmetry fits are made for different assumptions on the value of Δm_s, which is fixed in each of the fits. In each fit, the amplitude $\mathcal{A}$ is measured. For the true value of Δm_s, $\mathcal{A} = 1$ is expected, for all other values the amplitude should be zero.

2.3 B Flavour Tagging

The aim of flavour tagging is to determine whether a neutral B meson was produced as a particle or as an antiparticle. This section gives an overview on the physics of flavour tagging and existing approaches at the CDF experiment.

In figure 2.4, a schematic view of a typical B_s^0 event is given. At the primary vertex, where the initial reaction takes place, a $b\bar{b}$ quark pair is produced. Single b quarks can be created by electroweak processes, which have much lower cross sections than $b\bar{b}$ pair production within quantum chromodynamics. At the Tevatron, most b quarks are created in pairs, so the presence of a $b\bar{b}$ quark pair in the event is assumed in following.

In a B_s^0 event, at least one of the b quarks forms a B_s^0 meson. The corresponding hemisphere of the event is called the *same side*. The side of the remaining b quark is called the *opposite side*. This naming scheme is consequently applied to the B mesons or hadrons on both sides, for their decay vertices, tracks originating from them and for their jets. As $b\bar{b}$ quark pairs are not always produced back to back, a spatial overlap between same side and opposite side is quite common.

Flavour tagging approaches can be divided in *same side tagging* and *opposite side tagging*. Taggers that are exclusively focused on opposite side effects are independent of the actual type of B meson on the same side and can be used for B_s^0, B^0 and B^+ mesons without special adaptation.

2.3.1 Opposite Side Tagging

Opposite side tagging is based on the fact that the opposite side b quark has the opposite flavour of the same side b quark at production time. If the quark flavour on the opposite side can be determined, the production flavour on the same side is also known.

Unfortunately, neutral B mesons on the opposite side can also oscillate to their antiparticle. This can lead to a wrong tagging decision in methods that determine the opposite side flavour at decay time of the opposite side B hadron. Lepton and kaon tagging are affected by this problem.

Due to the variety of possible decays and limited detector acceptance it is usually not possible to fully reconstruct the opposite side B hadron, so inclusive properties need to be used.

Leptons

About 20% of B hadrons decay semileptonically, having a muon or an electron in the final state. The lepton charge is correlated to the flavour of the heavy quark: $b \rightarrow l^-$ and $\bar{b} \rightarrow l^+$, l being a muon or an electron.

In principle also τ leptons can appear here. As they are hardly reconstructed in the detector, only electrons and muons are used for tagging. A considerable number of τ leptons decay into a muon or an electron and two neutrinos. Leptons from these decays contribute to electron or muon tagging.

Besides mixing on the opposite side, cascade decays are another physical reason for wrong decisions in lepton tagging. There is a large branching fraction for the decays $b \rightarrow c \rightarrow l^+$ and $\bar{b} \rightarrow \bar{c} \rightarrow l^-$. If the lepton used for tagging originates from such a c quark decay, the sign will be swapped.

Kaons

Cascade decays, which are a limitation in lepton tagging, can be exploited in opposite side kaon tagging: According to the CKM matrix, the most probable decay path is $b \rightarrow c \rightarrow s$. The s quark will often end up in a kaon. If the kaon is charged, its charge is correlated to the opposite side b quark flavour.

Jet Charge

Jets are formed during fragmentation of quarks into hadrons. The charge of the entire jet is correlated to the charge of the initial quark. Jet charge tagging tries to determine the charge of the opposite side B decay jet by a weighted sum of its track charges. The identification of the right jet is a nontrivial task, and also the decision about which tracks belong to the selected jet is not clear without ambiguity.
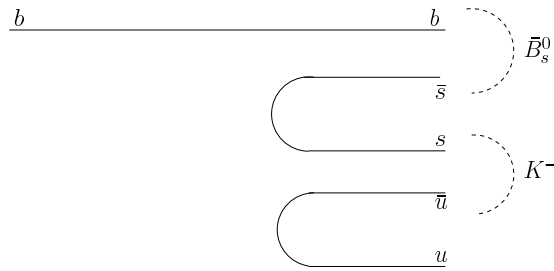
Figure 2.5: Fragmentation process of the same side b quark: The Kaon charge can be used to determine the B_s^0 state at production time.

At least one reconstructed jet on the opposite side is necessary for this method. The opposite side B hadron type and its decay are of minor importance. In about 95% of all events a jet charge tag can be made.

Vertex Charge

If the opposite side B hadron was charged and its decay vertex is known, the charge of the hadron can be determined by summing up charges of all tracks originating from that vertex.

Usually the opposite side decay vertex is not well known, so the quality of this tag is limited. Furthermore, decisions can only be made on charged hadrons, which reduces efficiency. There are no problems concerning mixing on the opposite side because charged mesons cannot oscillate to their antiparticle.

2.3.2 Same Side Tagging

The same side B meson emerges from the hadronization of the same side b quark. As the s or $\bar{s}$ quarks that are necessary to form B_s^0 mesons differ by their charge, the remaining fragmentation tracks have a charge bias correlated to the b quark charge. Same side flavour tagging is based on this effect.

To form a $\bar{B}_s^0$ meson from a b quark, an $\bar{s}$ quark is needed. It is obtained by pulling a $s\bar{s}$ quark pair from vacuum, so an s quark is left. This process is illustrated in figure 2.5. The probability that it will end up in a kaon is large. If this kaon is charged, its charge has the same sign as the b quark in the same side $\bar{B}_s^0$ meson.

The same principle is used for flavour tagging in the B^0 system. Instead of an $s\bar{s}$ quark pair, a $d\bar{d}$ pair is involved. The leading fragmentation partner of the B^0 meson is a pion instead of a kaon in the B_s^0 case.

2.3.3 Tagging Performance

To measure the performance of flavour tagging, the quantities *dilution*, *efficiency* and *tagging power* are used and will be explained in the following paragraphs.

Dilution

In a binary decision like flavour tagging, half of the answers are expected to be true in case of random decisions. If all decisions are wrong, one can correct that simply by swapping the sign of the decision. To measure the performance of a tagging algorithm it is therefore useful to define a quantity that is larger than zero if there are more right then wrong decisions, is zero if half of the decisions are right and smaller than zero if there are more wrong decisions. Furthermore is it convenient to normalize this quantity to the number of decisions made, so that its values are from the interval $(-1, 1)$. In flavour tagging terminology, this quantity is called *dilution*, which is misleading, because a high dilution is desirable for a flavour tagging algorithm. In the definition of the dilution

$$D = \frac{N_{RS} - N_{WS}}{N_{RS} + N_{WS}} \tag{2.29}$$

N_{RS} is the number of right and N_{WS} the number of wrong decisions made. It is of great use in the mixing fit to have an event by event prediction on the dilution, as it can be used to give events with a better tagging decision a higher weight in the fit for the asymmetry.

Efficiency

As many algorithms do not give a tag to every event (for example only to events containing leptons), the *efficiency* ϵ is defined as the fraction of events with a flavour tag:

$$\epsilon = \frac{N_{RS} + N_{WS}}{N_{RS} + N_{WS} + N_{NT}}, \tag{2.30}$$

where N_{NT} is the number of events where no tag is given at all. The flavour tagging algorithm described in this thesis is designed to give an answer for every event that has at least one track not used in the exclusive reconstruction. Its efficiency is therefore higher than 99%.

Tagging Power

As given by equation 2.23, the significance of the B_s^0 mixing analysis depends on $\sqrt{\epsilon D^2}$, so the *tagging power* $\mathcal{T}$ is defined as

$$\mathcal{T} = \epsilon D^2. \tag{2.31}$$
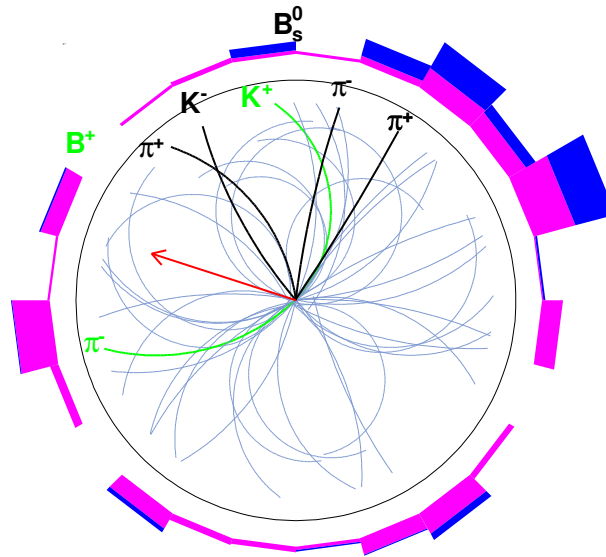
Figure 2.6: Event display of a B_s^0 decay in the channel $B_s^0 \to D_s^- \pi^+, D_s^- \to K^- \pi^+ \pi^-$ in transverse projection. The tracks of the same side B_s^0 decay are black. Some tracks of the opposite side B^+ decay (green) are missing, the red arrow denoting missing transverse momentum. Energy deposition in the calorimeter is indicated by the radial columns outside the black circle corresponding to the tracker. Background tracks are grey.

It corresponds to the size of a hypothetical sample with perfect tagging decisions that has the same overall statistical power as the reference sample.

2.3.4 Flavour Tagging at CDF Today

Challenges in Hadronic Environment

Due to the high total proton-antiproton cross section, events in a hadron collider like the Tevatron are usually full of background tracks. Besides the interesting process, in this case the production of a B_s^0 meson, there are tracks from proton remnants fragmentation and from multiple interactions per beam crossing. To illustrate this, part of an event display from the CDF detector is shown in figure 2.6.

The B decay tracks on both sides are not collimated to form the typical jet structure. As the $b\bar{b}$ quarks are not necessarily produced back to back, their decay regions are often overlapping, complicating separation of the two sides. Because of limited detector acceptance, there are often opposite side B tracks missing. These challenges render flavour tagging a difficult task in a hadronic environment.

Available Taggers

For the current measurement of B_s^0 oscillations [5] at CDF, a combination of several flavour taggers was used, each of them dealing with a separate tagging aspect. For opposite side tagging there are likelihood based soft lepton taggers for muons [20] and electrons [21], the opposite side kaon tagger [22] and the neural network based jet charge tagger [23]. An artificial neural network is used to combine all opposite side taggers [24], yielding a tagging power of

$$\mathcal{T}_{\text{NNOST}} = (1.8 \pm 0.1) \ \%. \tag{2.32}$$

Same side tagging is done by a same side kaon tagger [25] based on a neural network combination of a kaon identification likelihood with kinematic track quantities. Its tagging power

$$\mathcal{T}_{\text{SSKT}} = 3.7\% \tag{2.33}$$

in hadronic B_s^0 decay modes [5] is roughly twice as large as the power of the combined opposite side tagger.

Same side and opposite side tags are regarded as independent measurements of the b quark flavour. The predicted dilution D of both tags (D_1 and D_2) is combined to

$$D = \frac{D_1 + D_2}{1 + D_1 \cdot D_2} \tag{2.34}$$

when the tags have the same sign, or to

$$D = \frac{D_1 - D_2}{1 - D_1 \cdot D_2} \tag{2.35}$$

when the tags have different signs. The combined same side and opposite side tagging power is not stated, but it is certainly between the power of the same side tagger alone and the sum of both taggers:

$$3.7\% \leq \mathcal{T} \leq 5.5\%. \tag{2.36}$$

2.4 Experimental Status

In 2006, the CDF collaboration measured the B_s^0 oscillation frequency for the first time at a significance level of more than 5σ [5]. The measurement was based on a dataset of 1 fb^{-1} and made use of 5600 fully and 3100 partially reconstructed hadronic B_s^0 meson decays as well as 61500 partially reconstructed semileptonic B_s^0 decays. For initial state flavour tagging, the neural network combination of all opposite side taggers and the same side kaon tagger were used. The B_s^0 oscillation frequency was determined in an unbinned
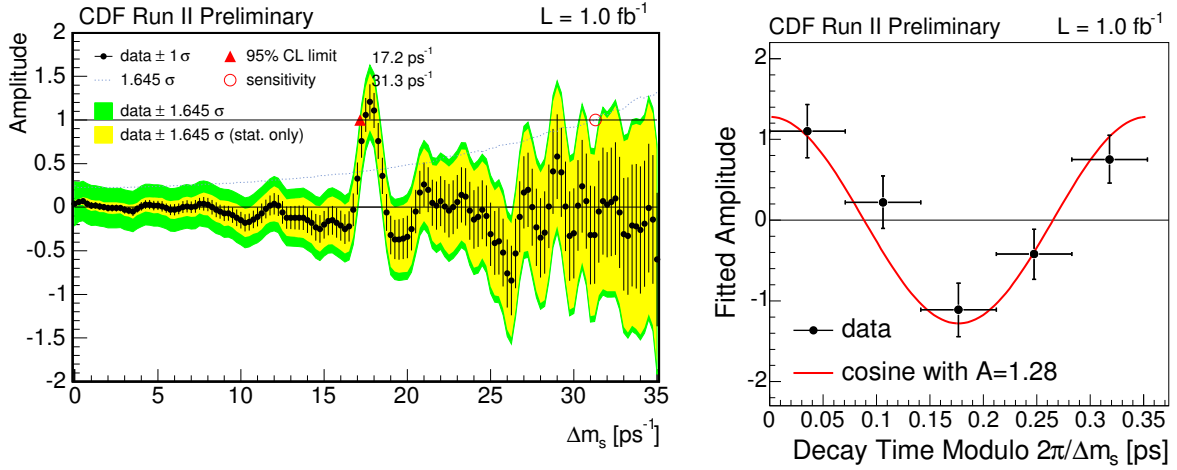
Figure 2.7: Result of the CDF B_s^0 mixing analysis [5]. The amplitude scan of all modes combined can be seen on the left. At the measured frequency the observed amplitude is incompatible with zero. The time dependent asymmetry with all later oscillations collapsed to the first period is shown on the right. It gives a visible impression of the oscillation signal.

maximum likelihood fit to be

$$\Delta m_s = 17.77 \pm 0.10 \text{ (stat.)} \pm 0.07 \text{ (syst.)ps}^{-1}. \tag{2.37}$$

The result of an amplitude scan as well as a visualization of the time dependent asymmetry can be seen in figure 2.7.

Since then, the available dataset has been increased by roughly a factor of five. An updated measurement will take advantage of the increased statistics and provide the only possibility for the validation of B_s^0 flavour tagging on data. Therefore a new analysis of B_s^0 oscillations is underway in the context of another PhD thesis within the Karlsruhe CDF group [26]. Preliminary results concerning the validation of the new flavour tagger are given in chapter 6.

Concerning CP violation in $B_s^0 \rightarrow J/\psi\,\phi$ decays, the latest CDF measurement was performed in summer 2008 on a dataset of 2.8 fb^{-1}, corresponding to 3200 signal events. The neural network combined opposite side tagger was used on the full dataset, while the use of the same side kaon tagger was limited to the first 1.35 fb^{-1}, because calibrations for particle identification were not available for later data. An unbinned maximum likelihood fit was performed to extract $\beta_s^{J/\psi\,\phi}$ and $\Delta\Gamma_s$. Due to undetermined strong phases, the likelihood functions shows two symmetric minima in the plane of the two parameters, which are overlapping and not of a Gaussian shape. Therefore no point estimates for $\beta_s^{J/\psi\,\phi}$ and $\Delta\Gamma_s$ can be given, instead the confidence regions have been evaluated using the
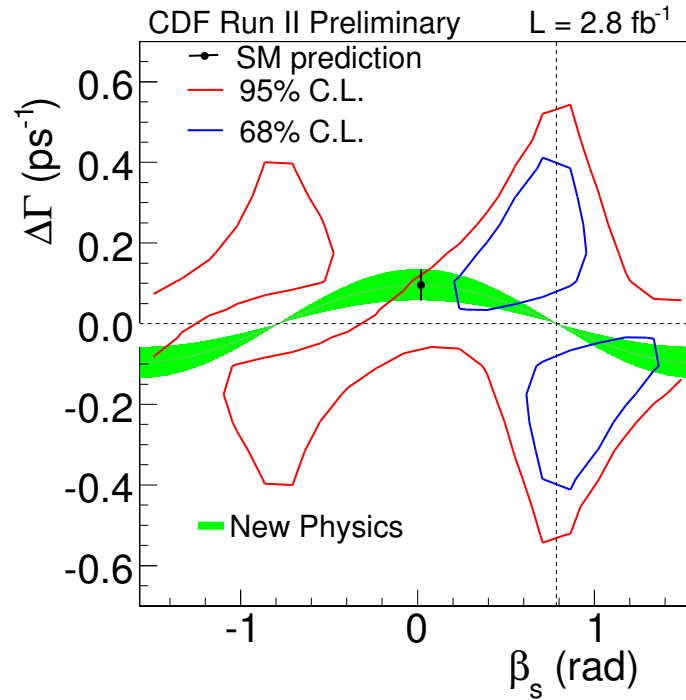
Figure 2.8: Result of the CDF analysis of the decay $B_s^0 \to J/\psi\,\phi$ to measure $\beta_s^{J/\psi\,\phi}$ and $\Delta\Gamma_s$. The two contours correspond to the 1σ and 2σ confidence regions. The Standard Model expectation is indicated by the black marker and is compatible with this measurement at a level of 1.8σ. The physical region under the assumption that new physics only affects M_{12} but not Γ_{12} is indicated by the green band.

Feldman Cousins method.

The resulting contours can be seen in figure 2.8, and the compatibility with the Standard Model expectation is 1.8σ. Also an update of this analysis is currently in preparation. As soon as the validation of the new flavour tagger has been fully established in the updated oscillation analysis, the $B_s^0 \to J/\psi\,\phi$ analysis will be the first place where it is used in the search for CP violation.

3 Experimental Setup

The Fermi National Accelerator Laboratory (*Fermilab*) is a high energy physics laboratory situated about 50 km west of Chicago in Batavia, Illinois (USA). It was founded in 1967 as National Accelerator Laboratory and was renamed in honour of Nobel Prize winner Enrico Fermi in 1974. Two Standard Model particles were discovered here, the b quark in 1977 and the top quark in 1995. Besides various activities in development of accelerator technology, fixed target and neutrino experiments, Fermilab is home of the *Tevatron* collider with its two experiments, DØ and CDF.

3.1 Accelerator

The Tevatron is a proton-antiproton-collider with a centre of mass energy $\sqrt{s} = 1.96 \,\mathrm{TeV}/c^2$. It is currently the highest energy artificial particle accelerator until the Large Hadron Collider at CERN successfully takes up operation. During *Run I*, the first period of operation between 1985 and 1996, a total luminosity of 110 pb^{-1} has been delivered at $\sqrt{s} = 1.8 \,\mathrm{TeV}/c^2$. Run II started in March 2001 and is expected to continue until at least summer 2010. Up to the 2009 summer shutdown, the Tevatron has delivered about 6.7 fb^{-1} to each of the two experiments.

Before the Tevatron can do its work, the two beams need to undergo several preparational steps, especially antiprotons need to be produced and conditioned. A schematic view of the accelerator complex necessary for this task is shown in figure 3.1 and will be explained below.

3.1.1 Proton Source

The Proton Source consists of the Preaccelerator, Linac and Booster [27]. In the Preaccelerator, hydrogen gas is first converted to negatively charged H^- ions and accelerated to a kinetic energy of 750 keV$/c^2$. The necessary static electric field of 750 kV is provided by a Cockcroft-Walton cascade generator.

After preacceleration, the H^- ions are passed on to the Linac via the 750 keV transfer line. The low energy side of the Linac consists of five RF drift tubes driven by a power amplifier tube each, while the high energy side has seven side coupled cavity linac modules powered by Klystrons. In between both sections, bunches are formed by the *vernier* and

FERMILAB'S ACCELERATOR CHAIN
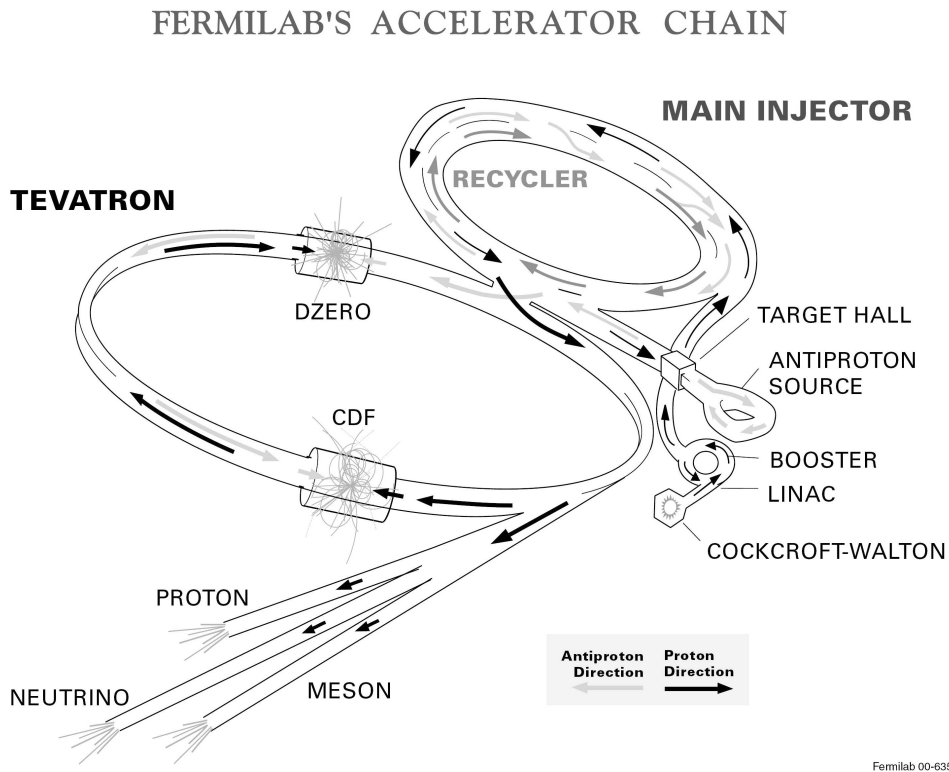
Figure 3.1: Overview of Fermilab accelerators, including antiproton production and preacceleration stages.

the *buncher*. At the end of the 150 m long Linac, the ions have been accelerated to 400 MeV/c^2.

Before the next stage of acceleration, a carbon foil target is used to strip the electrons from the H^- ions, leaving only bare protons. These are then accelerated to 8 GeV/c^2 in the Booster synchrotron, the first circular accelerator in the chain, with a radius of 75 m.

3.1.2 Main Injector

The Main Injector is a circular synchrotron seven times as large as the Booster and can accelerate or decelerate particles between the energies of 8 GeV/c^2 and 150 GeV/c^2 [28]. It is used for injecting beam into the Tevatron at 150 GeV/c^2, for antiproton production at 120 GeV/c^2 and for other uses in the various Fermilab experiments.

3.1.3 Antiproton Source

Unlike protons, antiprotons are not available in gas bottles. Production and conditioning of a suitable amount of antiprotons to fill the Tevatron is a challenging task and a limiting factor for the collider luminosity. The antiproton source consists of the target station, Debuncher and Accumulator. Also the Recycler plays an important role in antiproton handling.

A 120 GeV/c^2 proton beam from the Main Injector is directed onto a nickel alloy target, producing all sorts of secondary particles [27]. By using magnets selecting momentum and charge, antiprotons at 8 GeV/c^2 can be isolated from this particle spray.

The Debuncher, a rounded triangular-shaped synchrotron with an average radius of 90 m, captures the antiprotons by a special RF configuration called *bunch rotation*. To reduce beam emittance and momentum spread, stochastic cooling techniques [29, 30] are used here. On one side of the ring, a signal is picked up from the beam, amplified and applied to the same bunch on the other side of the ring, creating a negative feedback loop. Two transverse and one longitudinal cooling systems are available in the Debuncher.

The same triangular tunnel is also home of the Accumulator. It accepts antiprotons from the Debuncher and stores them at 8 GeV/c^2. This process is called *stacking*. The stack of antiprotons in the Accumulator is further cooled by several systems.

The final place where antiprotons are kept before injection is the Recycler. It is a storage ring located in the same tunnel as the main injector using, permanent magnets that maintain their field even when electrical power is lost. The original purpose of the Recycler was to collect the remaining antiprotons at the end of a Tevatron store instead of discarding them in the beam dumps. However, it turned out that this is not operationally feasible, as it would involve scraping all remaining protons away, decelerating the antiprotons and cooling them again. So despite its name, the Recycler is nowadays used as an extension to the Accumulator, which also greatly improved antiproton handling.

Besides stochastic cooling, the Recycler also uses *electron cooling*, which is a more efficient technique at higher beam intensities. When a certain amount of antiprotons is reached, a concentrated cool beam of electrons is injected on top of the antiprotons. A momentum transfer from the hot antiprotons to the cool electrons happens in the subsequent collisions, providing a substantial longitudinal cooling force.

3.1.4 Tevatron

The end of the accelerator chain is the Tevatron, a circular synchrotron with a circumference of 6.3 km. Both protons and antiprotons are kept in the same beam pipe, moving in opposite directions to allow head-on collisions. Collisions take place at two interaction points, where the CDF and DØ experiments are located.

The Tevatron is the only cryogenically cooled accelerator at Fermilab, and virtually all of its magnets are superconducting. They are made of niobium/titanium alloy and

need to be kept at temperatures below 4 K. Superconductivity allows for generating high magnetic fields without the danger of overheating from the high electric currents. The price for this advantage is that extensive cryogenic facilities and special magnet protection systems are needed for operations.

The Tevatron is mainly used as a storage ring. The stable situation of colliding 980 GeV/c^2 proton and antiproton beams is called a *store*. It starts with a *shot setup*, where the Main Injector delivers proton and antiproton bunches at 150 GeV/c^2 to the Tevatron, which are subsequently accelerated to the final energy within less than two minutes. After acceleration, collisions are initiated and the beam halo is removed by collimators. Then the store is ready for high energy physics use and lasts until it is lost or aborted to make room for a new one, typically after 16 to 20 hours.

The beam in the Tevatron consists of 36 bunches grouped in three *trains*, separated by *abort gaps*. When the beam needs to be aborted, abort kicker magnets are ramped up to guide it safely into the beam dump blocks. The abort gap is long enough to allow these kickers to ramp up to their full field.

3.2 The CDF Detector

The *Collider Detector at Fermilab* (CDF II) is a multi purpose solenoidal detector with subsystems for precise tracking, fast calorimetry and fine grained muon detection [31]. Figure 3.2 shows an elevation view of the CDF detector. The tracker is located inside a superconducting solenoidal coil, generating a 1.4 T magnetic field parallel to the beam axis. Calorimeters and muon systems are installed outside the magnet.

Following the symmetry of the detector, a polar coordinate system is used. The proton direction - towards the east at CDF - defines the z-axis. The polar angle θ is measured from positive z direction, while the pseudo-rapidity $\eta = -\ln(\tan(\theta/2))$ is commonly used when referring to that coordinate. The azimuthal angle ϕ is measured from the Tevatron plane, and r is defined in a plane orthogonal to the beamline.

3.2.1 Tracking System

Precise and efficient charged particle tracking is a major benefit in many CDF analyses [31]. It makes secondary vertex reconstruction possible, helps increasing the momentum resolution in meson reconstructions and is of great value in electron and muon identification.

The CDF tracker consists of two different detectors: A silicon tracker, formed by the subsystems *Silicon Vertex Detector* (SVX II), *Layer 00* (L00) and *Intermediate Silicon Layers* (ISL), and the *Central Outer Tracker* (COT), a cylindrical drift chamber. The arrangement of the tracking components and their respective η coverage is illustrated in figure 3.3.
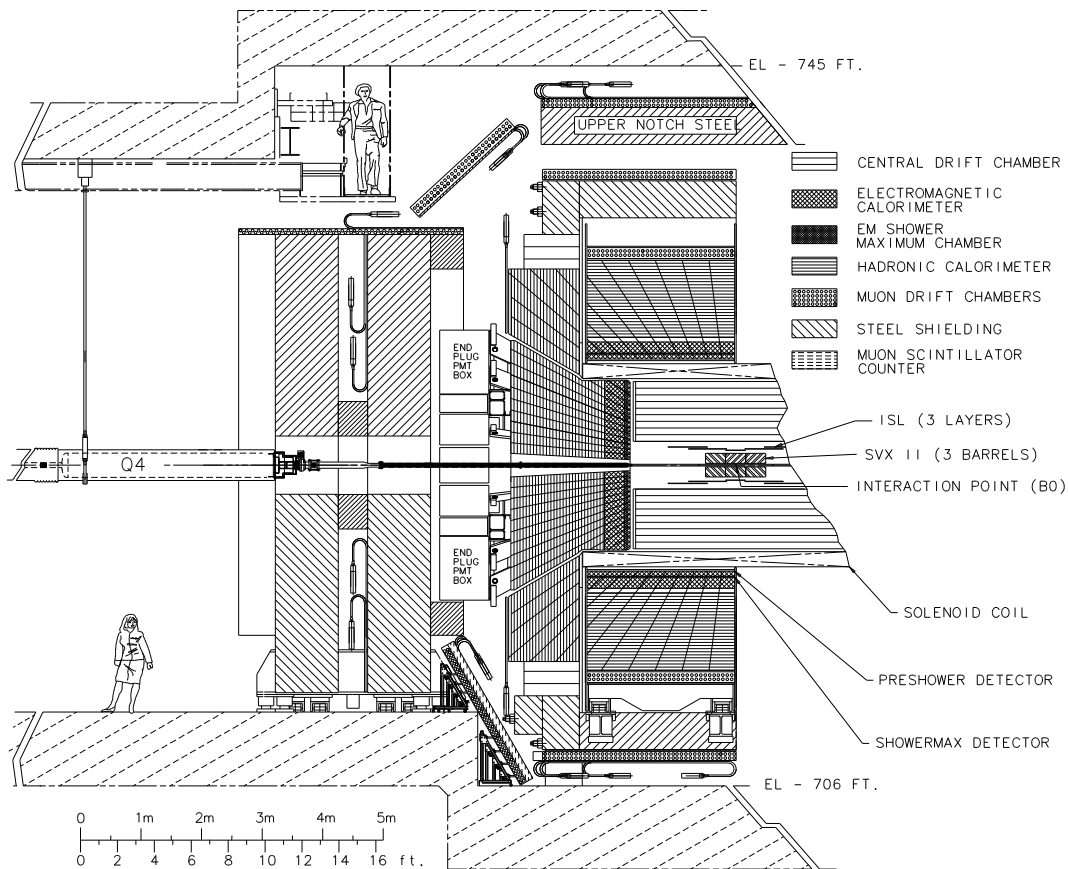
Figure 3.2: The CDF II detector.

Silicon Detector

Five layers of double sided silicon microstrip detectors form the SVX II, mounted in three barrels at a total length of roughly 1 m. On three of these layers, the strips on the second side are rotated by 90°, making precise z measurements possible, while the remaining two layers use a stereo angle of 1.2° for reducing ambiguities in case of multiple hits.

The L00 provides an additional layer of single sided silicon sensors mounted directly on the beam pipe, as close as possible to the interaction point. This late addition to the design significantly improves the impact parameter resolution and therefore vertexing capabilities.

The ISL consists of double sided sensors with a stereo angle of 1.2°, a single layer in the central region and two layers towards the detector end plugs. It is located between the SVX II and the drift chamber and helps matching tracks between both tracking subsystems.
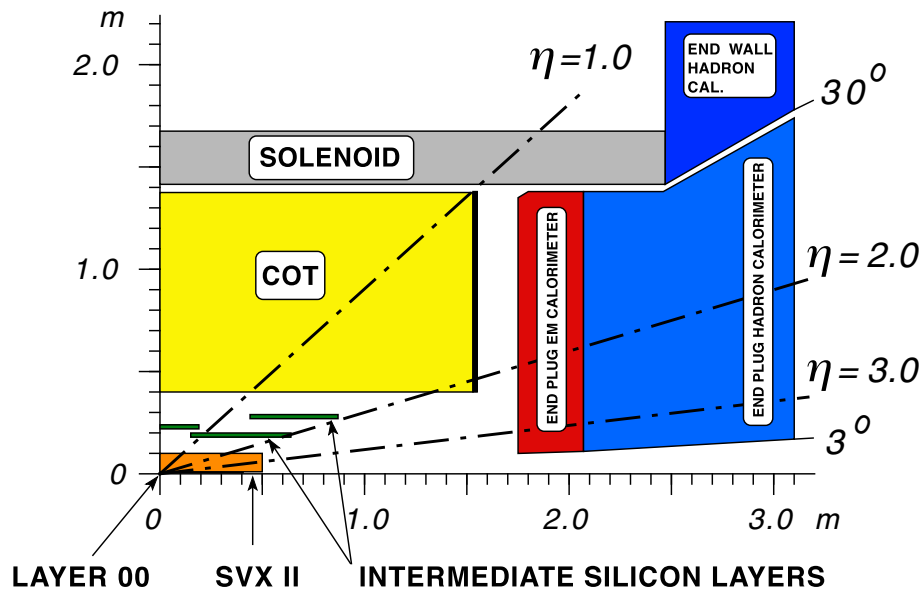
Figure 3.3: CDF tracking components and their η coverage.

Central Outer Tracker

The Central Outer Tracker [32] is a cylindrical drift chamber with a length of 3 m along the z axis, covering the radial region between 44 and 132 cm. Its 30240 sense wires are grouped in *supercells* of 12 wires each, which again are arranged in eight *superlayers*. Field sheets and potential wires are used to shape the electric drift field. A charged particle on its way through the COT interacts with the Ar-Et-CF$_4$ gas mixture, creating secondary charged particles. These are accelerated in the drift field, amplifying the signal by generating additional charged particles, and are detected as a charge deposition on the sense wires. The maximum drift distance is roughly 1 cm. Four *axial superlayers* have their wires parallel to the z direction, while the other four *stereo superlayers* have wires tilted by about two degrees, providing z position information. The maximum number of 96 hits for a given track allows a precise reconstruction of the transverse momentum p_T in the $r - \phi$ plane, which is complementary to the very good spatial resolution in the silicon detector close to the interaction point.

3.2.2 Particle Identification

Long-lived particles like protons, kaons and pions can not be distinguished by tracking information alone. In many physics analyses this is desirable, as for example kaons play an important role in flavour tagging. CDF allows particle identification based on two different physical sources: The specific energy loss in matter and the flight velocity.

The energy loss of a particle in matter is described by the Bethe-Bloch equation and depends on its velocity $\beta = v/c$ as well as on material properties. The energy deposition in the COT can be determined from the measured hit pulse length Δt, which is proportional to the logarithm of the deposited charge:

$$\frac{dE}{dx} \sim \ln(Q) \sim \Delta t. \tag{3.1}$$

Also the silicon detector could be used for measuring specific energy loss, which is not done because it would only yield a reasonable separation for very low energetic particles [33].

A velocity measurement is done with the *Time Of Flight* detector (TOF) [34], located between the COT and the solenoid. It consists of 216 scintillator bars roughly as long as the COT, leading to the same η coverage, and is used to determine the flight duration $t = t_{\text{TOF}} - t_0$ between the primary interaction t_0 and the arrival at the TOF detector t_{TOF}. In combination with the momentum measurement from the reconstructed track curvature, the relation

$$m = \frac{p}{c}\sqrt{\frac{(ct)^2}{L^2} - 1}, \tag{3.2}$$

where L is the length of the flight path, can be used to determine the particle's mass.

Figure 3.4 shows the momentum dependence of the performance of the two particle identification subsystems. It can be seen that both methods are complementary: Time Of Flight is a valuable separator at low track momenta, while dE/dx is more powerful in the higher momentum region.

3.2.3 Calorimeters

The CDF II calorimetry system consists of five calorimeter units [36] located outside the solenoid (see figure 3.2). Their purpose is to measure the energy of all particles except muons and neutrinos by full absorption. Electromagnetic [37] and hadronic [38] calorimeters are used in the central and the end-plug regions, with an additional hadronic-only end-wall calorimeter. The whole system covers the pseudorapidity region up to $|\eta| = 3.64$.

As the calorimetry system is optimized for higher track momentum regions, they are not of major importance for most B-physics analyses as the one described here. Their absorbers however form a valuable shielding to stop all other particles from reaching the muon chambers. More details on the calorimetry system can be found in the CDF technical design report [31].
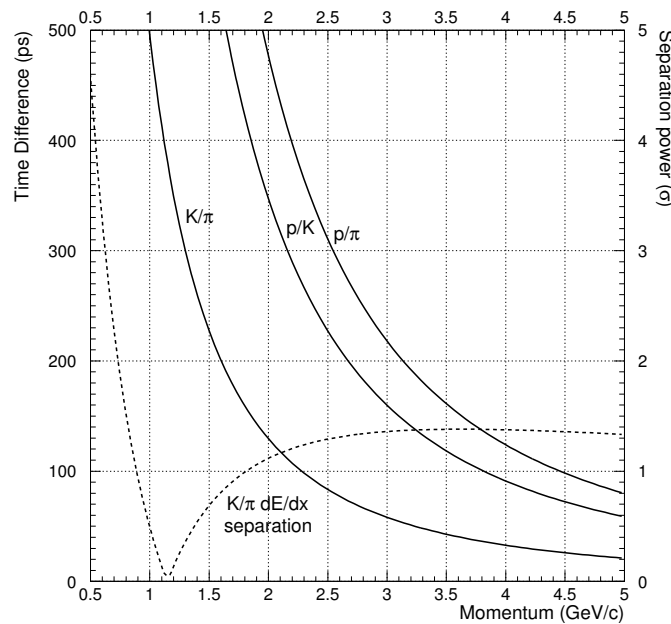
Figure 3.4: Performance of the particle identification system at CDF [35]. The solid lines show the separation power in the Time Of Flight system for different particle pairs, while the dashed line represents the kaon/pion separation by dE/dx in the COT. The typical time resolution of the Time Of Flight system is between 110ps and 130ps.

3.2.4 Muon Systems

Muons are more than 200 times heavier than electrons and do not take part in the strong interaction. Therefore they do not lose significant amounts of energy in both the electromagnetic and the hadronic calorimeter and cross them nearly undisturbed. As all other charged particles are absorbed there, the best muon identification is achieved in the outermost region of a detector.

The CDF muon system consists of four subdetectors [39, 40]. The central region up to $|\eta| < 0.6$ is covered by the *Central Muon Detector* (CMU) and the *Central Muon Upgrade* (CMP). The CMP is located outside extra shielding to increase the muon purity and has therefore a higher sensitivity threshold in transverse track momentum. The *Central Muon Extension* (CMX) covers the region $0.6 < |\eta| < 1.0$ and is tilted about $45°$ to the z axis. For mechanical reasons the modules at the bottom of the detector have to be smaller, slightly reducing the coverage in this ϕ-region. In the forward region, the *Intermediate Muon System* (IMU) extends the azimuthal coverage to $|\eta| < 1.5$. All muon

systems are basically small tracking chambers with sense wires in drift tubes, which are not fast enough for trigger use. Therefore all muon chambers except CMU are covered with fast scintillators of lower granularity to provide fast response that can be used in muon triggers. Also cosmic muons can be rejected by the timing information provided by the muon scintillators.

3.2.5 Trigger

Every hadron collider has a much higher interaction rate compared to an electron machine, but most interactions do not contain physics processes of interest. The b production cross section at Tevatron energy is about three orders of magnitude lower than the total cross section, so roughly one out of thousand collisions is interesting for B-physics. This ratio is even worse for other interesting physics processes such as top quark or $W^{\pm}$ and Z^0 boson production.

The Tevatron runs at a high collision rate of 1.7 MHz, and storing the detector readout for one event takes about 250 kB. Putting aside that some detector systems are too slow for such a frequent readout, just the pure amount of one Terabyte every two seconds makes it impossible to store all information. An online system for preselecting interesting events is needed for reducing the amount of data that is read out. CDF uses a sophisticated three-level trigger system for this task [41]. An schematic view of the data flow is given in figure 3.5.

Level 1

The first trigger level is a synchronous trigger entirely implemented in hardware. During the roughly 5 μs that are needed for the decision, the digitized detector information is kept in a 42 clock cycle pipeline. It uses information from all components except the silicon detector. Approximate tracking information is provided by the *Extremely Fast Tracker* (XFT) [42] already at this early stage, which compares precalculated tracking patterns to COT information in order to find tracks and obtain their p_T and ϕ information. Also muon and calorimeter information is available at this stage.

Level 2

After being accepted by Level 1, an event is passed to the second trigger stage [43]. Level 2 uses a reconfigurable software implementation running on special hardware. Because its decisions are allowed to take more time than Level 1, four events can be kept in buffers in the meantime. The additional time is used for refining the L1 decisions with finer detector granularity and better precisions. For finding secondary vertices, XFT information from L1 is combined here with silicon tracking information in the *Secondary Vertex Trigger* (SVT) [44]. This offers the unique opportunity for B-physics to enrich events containing

Dataflow of CDF "Deadtimeless"
Trigger and DAQ

Detector   **7.6 MHz Crossing rate**
            132 ns clock cycle

L1 Storage   **Level1:**
Pipeline: L1 trigger   **7.6 MHz Synchronous pipeline**
42 Clock   **5544ns latency**
Cycles Deep   **<50 kHz Accept rate**

            L1 Accept

L2 Buffers: **Level 2:**
4 Events   L2 trigger   **Asynchronous 2 stage pipeline**
                                         ~20μs latency
            L2 Accept   **300 Hz Accept Rate**

                                    L1+L2 rejection: 20,000:1

DAQ Buffers

            L3 Farm

            Mass
            Storage   PJW 10/28/96

Figure 3.5: Overview of the CDF trigger and data acquisition system. The 7.6 MHz crossing rate were evaluated as a possible mode of Tevatron operation in the design phase; today the crossing rate is about 1.7 MHz.

secondary vertices already at an early trigger stage. The raw event rate has been reduced by a factor 20000 after L1 and L2 trigger.

Level 3

The third trigger stage is a computer farm running similar software to what is used during the offline reconstruction, but with different parameters. It has therefore the best momentum and impact parameter resolution of all trigger levels. When an event is accepted by L3, it is sent to the Feynman Computing Center for final storage and offline reconstruction. The output rate of the L3 trigger is around 100 Hz, which corresponds to accepting one in 120000 events.

4 Classification with Artificial Neural Networks

This chapter gives an introduction to statistical classification using artificial neural networks. First performance measurement of a classification method is discussed, followed by a brief description of conventional classification methods. After a short excursion into biology to provide some analogies, the concept and implementation of artificial neural networks is described.

4.1 Statistical Classification

In high energy physics it is often necessary to classify data into categories, often called *signal* (data that are interesting for the particular analysis) and *background*, based on several measured variables. Several statistical classification methods exist, one of them being *artificial neural networks*, which are described in the next sections.

4.1.1 Performance Measurement

When making binary decisions between signal and background classes, there are two types of errors: Accepting background events as signal and rejecting signal events as background. To quantify them, purity

$$P = \frac{N_{\text{signal, selected}}}{N_{\text{signal, selected}} + N_{\text{background, selected}}} \tag{4.1}$$

and efficiency

$$\epsilon = \frac{N_{\text{signal, selected}}}{N_{\text{signal}}} \tag{4.2}$$

are defined, both ranging from zero to one. For the first kind of error, the quality of the signal sample is tainted by background events, so the purity is decreased. A purity of one means that all selected events are actually signal events. For the other kind of error, signal events are lost, lowering the efficiency of the selection. If all signal events have been selected, the method has an efficiency of one. The efficiency is sometimes referred to as *signal efficiency*.
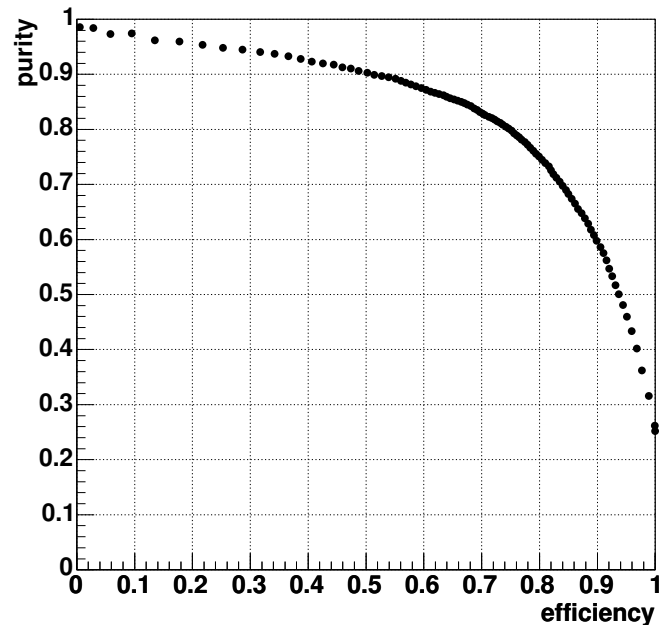
Figure 4.1: Example of a diagram showing purity as a function of efficiency of the identification of leptons from a B hadron decay. For many application, the cut that gets closest to the ideal point in the upper right corner is of interest.

Many classification methods provide information on the quality of their decision. In the ideal case, this is a scalar value that can be interpreted as a probability that the current event is signal or background. The actual classification is then done by selecting a single cut on such a quantity.

Every possible cut can be characterized by purity and efficiency. The performance of a method can therefore be illustrated by plotting purity as a function of efficiency for different cuts. An example of a purity/efficiency diagram is shown in figure 4.1. The ideal point is (1/1), which corresponds to having selected all signal events without accepting a single background event. The purity/efficiency pair of the selected cut is called the *working point*.

4.1.2 Conventional Approaches to Classification

A classification technique widely used is the method of *sequential cuts*. By looking at the characteristics of signal and background distributions of single variables, the best discriminating quantity is identified and a suitable cut applied. After cutting on the most important quantity, the procedure is repeated for the remaining variables.

Unfortunately, within the method of sequential cuts it is very difficult to take correlations between quantities into account. If there is a class of signal events that cannot be
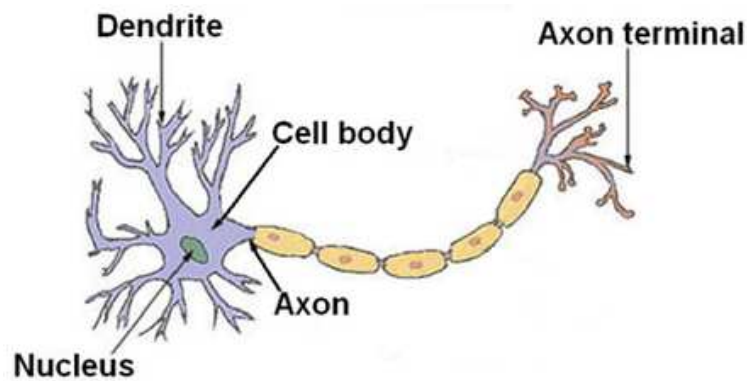
Figure 4.2: Structure of a typical neuron.

identified by looking at single variables, but is clearly visible in a combination of two of them, a cut based classification technique might fail. Although usually a high purity can be obtained, the efficiency tends to be lower. Selecting a specific working point is difficult.

It is useful to combine all variables available into one scalar variable called *test statistics* and find a decision by cutting on that. This can be done for example in a *likelihood ratio*. For all input quantities, a parametrization of their signal fractions is derived. Then the test statistics is composed by multiplying the current value of this parametrization for all measured input quantities. This is a good approach if all input quantities are independent. Correlations lead to overestimating the influence of distinguishing features found in more than one input quantity.

4.2 Neurons in Nature

The human brain has a lot of features that are very helpful for classification: It is fault tolerant, robust against inconsistent input data and can adapt itself quickly to various problems. Therefore it might be helpful to have a look at its architecture and try to implement a simple model in a computer program.

The smallest unit of the brain is a *neuron* (figure 4.2), a special type of cell that also forms our nerves. It is connected to other neurons by *dendrite* for incoming and *axons* for outgoing signals. Dendrite and axons of different neurons are connected by *synaptic junctions*.

Signal transmission between neurons is a chemical process resulting in a change of the electrical potential in the receiving cell. If a threshold value is reached, the receiving cell emits a signal of a certain strength and duration to all cells connected to its axons. A
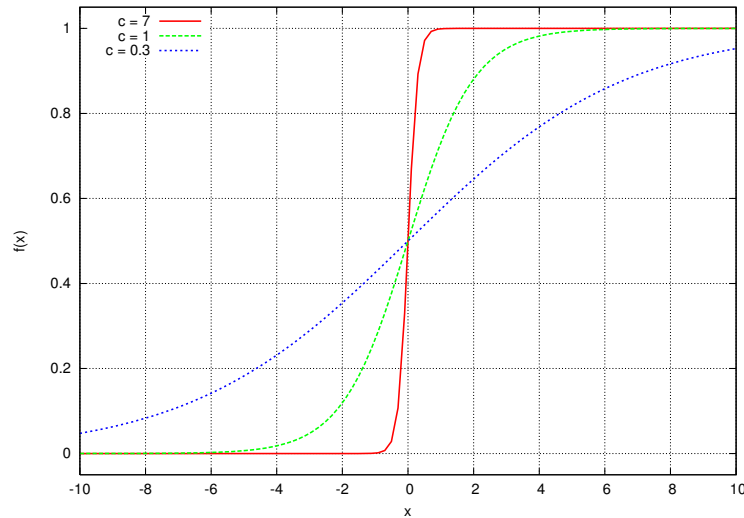
Figure 4.3: Sigmoid functions for different values of the parameter c.

simplified mathematical model of such a neuron n can be written as

$$n(t + 1) = \theta \left(\sum_i w_i n_i(t) - \mu \right) \tag{4.3}$$

with μ being the threshold value of the current neuron n, w_i representing the strength of the synaptic junction from neuron n_i and t being an index for a specific neuron in the chain. The transfer function θ is a step function in this simple model.

The brain is adapted to new tasks by changing the weights of the connections between neurons, adding new synaptic connections or *pruning* them away completely. The acquired knowledge of a neural network is stored in the weights of its connections.

4.3 Artificial Neural Networks

To construct an artificial neural network, the same basic concept of nodes (corresponding to neurons) and weighted connections is used. Instead of the simple step function from above, a *sigmoid* or *fermi* function (see figure 4.3) is used:

$$f(x) = \frac{1}{1 + e^{-cx}}. \tag{4.4}$$

Such a nonlinear function enables the neural network to learn nonlinear correlations between input variables. It maps the interval $(-\infty, \infty)$ to $(-1, 1)$ and can approximate a
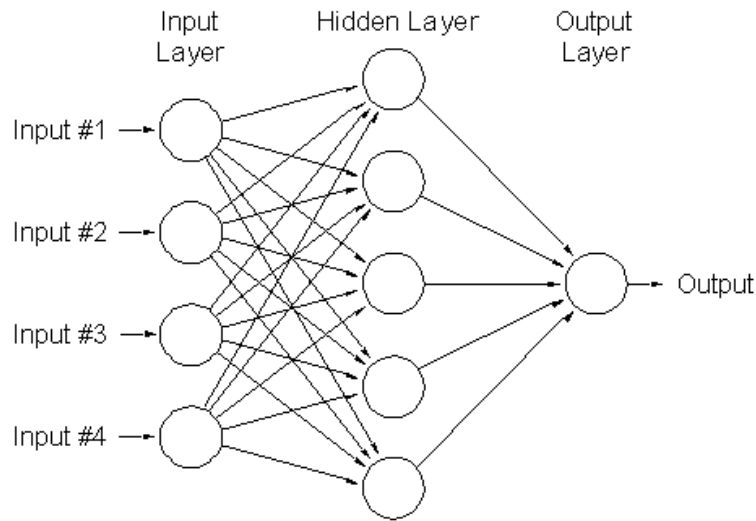
Figure 4.4: Multi layer feed forward neural network.

linear or a step function by variation of the parameter c.

There are several ways of grouping nodes into a network. A so called *single layer perceptron* consists of one layer of input nodes that are all directly connected to an output node. Neural networks used in classification usually have an additional *hidden layer* between input and output nodes. As there are no connections looping back to previous layers, these neural networks are also referred to as *feed forward* networks. A typical topology of such a neural network can be seen in figure 4.4.

For other applications as for example pattern recognition, a more complex network structure may be useful, including loops to previous neurons or even to the same neuron itself. These architectures will not be discussed here.

4.4 Neural Network Training

As mentioned above, all experience of a human brain is stored in the weights of its synaptic connections. Therefore, training an artificial neural network means finding a set of weights which maximizes knowledge.

A data sample is needed for training where the answer of the classification is already known, called *training sample*. This can be historical data, simulated events or sometimes even real data from the same measurement as the data to classify later. After training, the neural network is applied to the *classification sample*.

In a classification task, the neural network should be able to recognize important features of signal and background events from the training sample and base its decision upon them. There is a danger of *overtraining* the neural network. In that case, statistical fluc-

tuations of the training sample are 'learned by heart'. This can be detected by testing the neural network on a part of the training sample left out from the training process.

Neural network training is based on the minimization of a loss function $E(w)$ rating the overall difference between the neural network output $O(w)$ for the current weights w and the true value T. A χ^2 loss function is one of the most popular:

$$E(w) = \sum_{\mu} \left(O_{\mu}(w) - T_{\mu} \right)^2 \tag{4.5}$$

where μ is running over all training events. The loss function has a global minimum for $O_{\mu} = T_{\mu}$. Basically every differentiable function with this property can be used as a loss function, as for example an *entropy* based definition, which is used for neural networks in this thesis:

$$E(w) = \sum_{\mu} \log \left(\frac{1 + O_{\mu} \cdot T_{\mu}}{2} \right). \tag{4.6}$$

As shown in [45], the neural network output can be interpreted as a probability if the global minimum of the loss function is found.

4.5 NeuroBayes

Although the basic principle of an artificial neural network is simple and easy to understand, its implementation is far more difficult. Deep statistical knowledge and experience in numeric programming is necessary. It it therefore useful to rely on a software package that is carefully designed and thoroughly tested.

For all neural networks in this thesis, the *NeuroBayes* [45] package is used. NeuroBayes was initially written by Michael Feindt and other members of the *Institut für Experimentelle Kernphysik* of the *Karlsruhe Institute of Technology* and makes a lot of expert knowledge easily available to the user. Since 2002 it is further developed, promoted and applied to data analysis tasks outside high energy physics by Phi-T GmbH.

4.5.1 Preprocessing

NeuroBayes offers several preprocessing possibilities for input data that can be used to suppress statistical fluctuations and to evaluate correlations among variables or to the training target. The individual variable preprocessing options that were used in the neural networks within this thesis are explained here, along with a description on how to read the preprocessing plots shown at various places in the other chapters.

Continuous variables can be distributed in various ways and across wide number ranges, which can lead to numerical complications when used in their raw form. Interesting information is often compressed to a very small range of numbers in peaking regions,
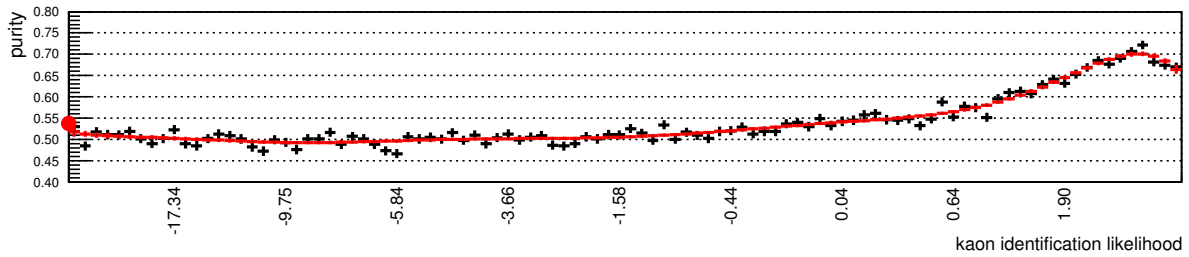
Figure 4.5: Example of individual variable preprocessing in NeuroBayes. Each of the 100 bins contains one percent of all input events, covering non-equidistant regions of the original input quantity. The fraction of signal events in each bin is denoted by the black data points. The red fit function is a regularized spline, suppressing statistical fluctuation of the data points. The overall signal fraction of the whole training sample is indicated by the red point on the y axis.

while comparably uniform behaviour in tail regions often evens extends across more than an order of magnitude. The first preprocessing measure is therefore usually to transform the input quantity into a flat distribution. This is achieved by means of a non-equidistant binning, leading to 100 bins of equal statistics. That technique is called *importance sampling* and also helps to suppress the influence of random fluctuations in low-statistics regions of the original distribution. Figure 4.5 gives an example of a preprocessing plot which can be used to follow the single steps that are described here.

As a next step, the fraction of signal events in each bin, called *purity*, is determined. To further suppress random noise, the purity as a function of the bin number is parametrized by a regularized spline fit. The example in figure 4.5 clearly shows that the fit describes the overall shape very well while effectively smoothing out fluctuations of the data points around it. Also a monotonous spline fit can be chosen by the user if warranted by the shape of the purity to be fitted. A similar procedure exists for input data consisting of discrete values where the spline fit is replaced by a regularization algorithm. For all further steps, the function value of this spline fit is used for representing the corresponding input quantity in a given event. At this point, the transformed variable can be interpreted as signal probability.

After the individual variables have been preprocessed, correlations between them as well as their correlations to the training target are determined. Then a principal component analysis in the space of all input variables is performed, effectively removing linear correlations between them. This set of input quantities in their new decorrelated base is then used to train the artificial neural network. Evaluating the linear correlation separately, the neural network is free to learn higher order correlations between variables without being distracted by information that can be handled by analytic procedures before.

As an alternative to a network training, one more transformation of the variable set

can be performed to project all target correlation into a single quantity. The purity of this quantity is parametrized by a monotonous spline fit similar to individual variable preprocessing and can then be used instead of the actual output of the neural network. In many cases the separation power of this quantity is equal to that of the neural network result.

5 Combined Flavour Tagging Using Neural Networks

The objective of the work presented in this thesis is to provide a general flavour tagging tool that makes use of all information available in an event. It should take advantage of correlations between the tagging information sources and combine all the information into a single probability variable for the flavour of the quark that is to be tagged.

Existing flavour tagging algorithms at CDF focus on one information source, for example on opposite side leptons. If there is no lepton, the given event is not tagged. In contrast to this, our tagger provides a decision for every event. To achieve this ambitious goal, advanced classification techniques are necessary. Artificial neural networks are used throughout all classification tasks arising on the way, and wherever possible the application of hard cuts is avoided. Instead, the probability delivered by one neural network is used as an input into the next stage.

Existing tools are used where they are available, for example NeuroBayes for all neural networks and the LeptonTools package [46] for lepton identification. The basic idea of this tagger is inspired by the BSAURUS [47] package from the DELPHI experiment. Its implementation is based on the jet charge tagger [23] and was started in preceding work [48, 49]. It has been further developed on Monte Carlo and validated on real data.

The flavour tagger has been developed for application in the B_s^0 meson system. However, from the B_s^0 version two other taggers for B^0 and B^+ mesons can easily be derived by training the relevant neural networks on a corresponding Monte Carlo simulation for another B meson family. These additional taggers provide valuable cross checks, as it is much easier to evaluate their performance on data than for the B_s^0 case. The variables that are regarded useful by the neural network training differ for the three cases, and flavour correlations are also not expected to be identical. In this description of the tagging strategy, the B_s^0 case is described. Information on the other two taggers will be given along with the description of calibration and performance evaluation in the next chapter.

5.1 Strategy

The architecture of this flavour tagger can be divided into three major parts:

- Preparing quantities that are useful for the later stages,

- examining every single track for its flavour information,

- combining everything into an event level flavour tag.

In the first step, an inclusive reconstruction of the opposite side B decay vertex is done - an exclusive reconstruction would not be feasible due to the variety of possible B hadrons and decay channels as well as limited detector acceptance. Spatial and kinematical relations between other tracks and the vertex are subsequently calculated. Combining this vertex information with electron and muon identification, leptons from the opposite side B decay are identified for later use in tagging.

The second step is flavour tagging on track level. Every charged track can either have the same or the opposite charge sign as the same side b quark. A probability value for this *track flavour correlation* is constructed. Depending on the region in the event the flavour correlations can differ a lot. Therefore the tracks are divided into natural subsamples: Tracks in the same side jet, lepton candidate tracks and the other remaining tracks. Each sample is then handled separately. Most tracks do not carry much flavour information, but as there are several of them in every event, the combination of their small correlations yields significantly more power than every track alone.

Finally, the track flavour correlation is combined to an *event level flavour tag* in one probability ratio per subsample and one combined probability ratio using all tracks. These probability ratios and an estimation of the opposite side vertex charge are used in a final neural network to make a tagging decision on event level. The following sections will describe each of the steps in more detail.

5.2 Variable Definitions

Several quantities on three different levels are used in the followings steps. The base of the variable hierarchy is the collision event in which a B meson candidate was reconstructed, defining the *event level*. Every event consists of at least the tracks used in the reconstruction of the B meson. Information about single tracks is defined on *track level*. It will be explained in the next section that up to six secondary vertex candidates for a second B hadron decay are formed in every event. Quantities describing these vertex candidates are available on *vertex level*. All variables that are relevant for the tagging procedure or the validation process are described here to give the reader a central reference point.

5.2.1 Event Quantities

The following quantities are defined once for every reconstructed same side B meson:

- *run*: The CDF run number in which a given event was reconstructed. Every run of the CDF data acquisition system has a unique run number assigned.

- *event*: The event number, also known as the trigger number, in which the given B meson candidate was reconstructed. Together with the run number it can be used to uniquely identify a collision event.

- *trigger*: The trigger scenario in which the event was recorded.

- *charge*: The charge of the B meson.

- *flavour*: The flavour of the B meson at decay time, determined from the charge of its decay products.

- *mass*: B meson invariant mass.

- *transverse momentum p_T*: The momentum component in the $r - \phi$ plane.

- *pseudo-rapidity η*: The pseudo-rapidity of the reconstructed B meson.

- *bMesonProb*: The output of a neural network for distinguishing B mesons in the given decay channel from combinatorial background.

5.2.2 Track Quantities

Track description quantities are grouped into categories: Standard track parameters, relation to the jet, to the same side B meson and to the opposite side secondary vertex.

Standard Track Information

Standard track parameters are provided by the CDF tracking code and used in many B physics analyses. When a spatial reference point is necessary, the origin of the detector coordinate system is used unless stated otherwise. The *perigee* is the point on the track helix that is closest to the reference point in the $r - \phi$ projection.

- *charge*: The track charge, measured from the direction of the track curvature in the magnetic field of the tracker.

- *transverse momentum p_T*: The projection of the track momentum to the $r - \phi$ plane.

- *pseudo-rapidity η*: The pseudo-rapidity of the track, where θ is the polar angle at the perigee (see also section 3.2).

- *ϕ*: The azimuthal angle ϕ at the perigee.

- *impact parameter d_0*: The minimum distance in the $r - \phi$ plane between the track helix and the reference point.

- z_0: The distance along the z axis between the track perigee and the reference point.

Besides the track parametrization itself, also information about the track quality is available:

- *nSIHits*: The number of hits in the silicon detector that are associated with the track. It can be further distinguished between hits with axial and stereo information.

- *nSIHitsL00*: The number of hits in Layer 00 of the silicon detector.

- *nCTHits*: The number of hits in the COT that are associated with the track. This can be subdivided into axial and stereo hits.

- *fit χ^2*: The χ^2 value of the track fit.

Relation between Track and Same Side B Meson

These track variables relate a track to the same side B meson:

- *signed impact parameter d_0*: The signed impact parameter as explained in figure 5.1, using the direction of the same side B to define the sign.

- $\Delta\eta$: The difference between the pseudo-rapidity of the track and of the same side B meson.

- $\Delta\phi$: The difference between ϕ of the track and of the same side B meson.

- ΔR: The spatial difference $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ between the track and the same side B meson. Also the rank in ΔR is available, with the first ranks corresponding to the smallest values of ΔR.

- $p_L^{rel(B)}$: The projection of the track momentum to the same side B meson direction. Also the rank in $p_L^{\rm rel}$ starting with high momentum tracks.

- $p_T^{rel(B)}$: The track momentum component perpendicular to the momentum of the same side B meson. The rank in $p_T^{\rm rel}$ starts at low momentum values.

- *invariant mass*: The invariant mass of the combination of the track with the same side B meson, assuming pion mass for the track. The rank in this quantity starts at low mass values.
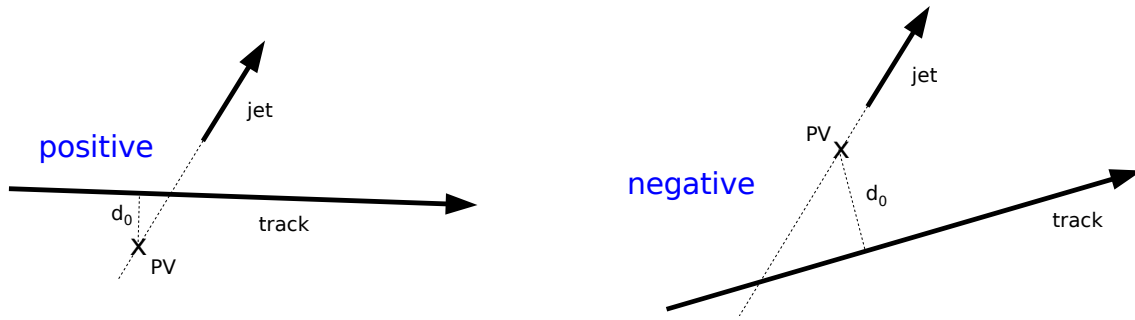
Figure 5.1: Definition of a signed impact parameter d_0, in this case related to a jet. If the intersection between the track and the jet axis is on the same side of the primary vertex (PV) as the jet direction (left illustration), the impact parameter is given a positive sign. If it is behind the primary vertex (right illustration), the sign is negative. Apart from a jet, any momentum vector can be used as a reference, for example a reconstructed particle.

Relation between Track and Jet

It will be explained below in subsection 5.3.3 that tracks are grouped into jets by a cone clustering algorithm. For an individual track, the following variables describe the relation to the jet it belongs to:

- *inSameSideJet*: A flag if the same side B meson was reconstructed in the same jet as the current track.

- *jet signed impact parameter* d_0^{jet}: The signed impact parameter as illustrated in figure 5.1 with respect to the jet direction.

- $p_T^{rel(jet)}$: The track momentum component perpendicular to the jet momentum.

- *rapidity*: The rapidity along the jet axis, assuming pion mass. Also the rank in this quantity is available, starting with the highest rapidity.

- p_T^{jet}: The transverse momentum of the jet associated with this track. It has the same value for all tracks that form a given jet.

- *nTracksJet*: Number of tracks in the jet associated with this track. It has the same value for all tracks that form a given jet.

Relation between Track and Opposite Side Secondary Vertex

The relation of a track to the opposite side secondary vertex is described by the following quantities:

- *usedInFit*: A flag which is true when the current track was used in the fit of the opposite side vertex candidate.

- *inJet*: A flag which is true if the track is in the same jet as the tracks used for the secondary vertex candidate.

- *mass*: The invariant mass of the track combined with the vertex, assuming pion mass for all tracks.

- p_{rest}: The track momentum transformed to the rest frame of the opposite side B decay vertex.

- $\cos\theta_{rest}$: the cosine of the helicity angle θ between the track in the opposite side secondary vertex rest frame and the corresponding opposite side B direction in the laboratory frame $\cos(\theta_{\mathrm{rest}})$

- *impact parameter* d_0: The impact parameter as defined in standard track data, but using the opposite side secondary vertex as reference point. Also the error on the impact parameter σ_{d_0} is available.

- z_0: The z difference as defined in standard track data, but using the opposite side secondary vertex as reference point. Also the error σ_{z_0} is available.

5.2.3 Secondary Vertex Quantities

Secondary vertex candidates are fitted from track pairs that have large probabilities to originate from the opposite side B decay. This process is described below in section 5.3.4, also the explanation of the *B track networks* and the *secondary vertex network* is given there. The properties of a secondary vertex candidate are described by the following variables:

- *bProb1*: The output of the first B track network for the first of the two tracks. The first track is always the one with the higher probability to originate from a B decay. Also the rank in this quantity is provided.

- *bProb2*: The output of the first B track network for the second of the two tracks. The second track is always the one with the lower probability to originate from a B decay. Also the rank in this quantity is provided.

- *inSameJet*: A flag which is true if both tracks forming the vertex candidate are associated with the same jet.

- *charge*: The sum of the two track charges, can be either 0 or ± 2.

- *mass*: The invariant mass of the two track combination, assuming pion mass for both tracks.

- L_{xy}: The distance of the vertex to the primary vertex in the xy-plane. Also the error $\sigma_{L_{xy}}$ is provided.

- $\Delta\phi$: The difference in ϕ between the vertex momentum and the vector pointing from the primary to the secondary vertex.

- $\Delta\eta$: The difference in η between the vertex momentum and the vector pointing from the primary to the secondary vertex.

- *fit* χ^2: The χ^2 value of the vertex fit and the corresponding fit probability.

5.3 Opposite Side Vertexing and Other Prerequisites

5.3.1 Monte Carlo Dataset

For a neural network training, simulated events with similar properties as measured events are needed. To achieve a good agreement between data and Monte Carlo events, some aspects of the simulation process need special attention. The procedure is described for the B_s^0 case, but equally applied to the B^+ and B^0 cases.

Using the event generator *Pythia* [50], $q\bar{q}$ events with all quark flavours are produced and immediately discarded if they don't contain b quarks. When Pythia is asked to explicitly generate $b\bar{b}$ events, only the leading order process *flavour creation* is simulated. The method we use also includes $b\bar{b}$ quark pairs from *flavour excitation* and *gluon splitting*, which is necessary for a proper description of spatial and kinematic relations of the two b quarks [23].

The decay of the hadron containing b and c quarks is simulated by EvtGen [51]. For each event containing at least one B_s^0 meson, exactly one B_s^0 meson is forced into the decay channel $B_s^0 \to D_s^{\pm}\pi^{\mp}$. All events without a B_s^0 meson are removed. Finally, a simulation of the full CDF detector and trigger is done. Events surviving the two track trigger are reconstructed, resulting in a dataset of roughly one million events.

5.3.2 Track Selection

All tracks that are used in this flavour tagger need to pass the following requirements:

- The track was not used in the exclusive reconstruction of the B meson to be tagged (the *same side B meson*).

- The transverse momentum of the track is at least 400 MeV/c^2.
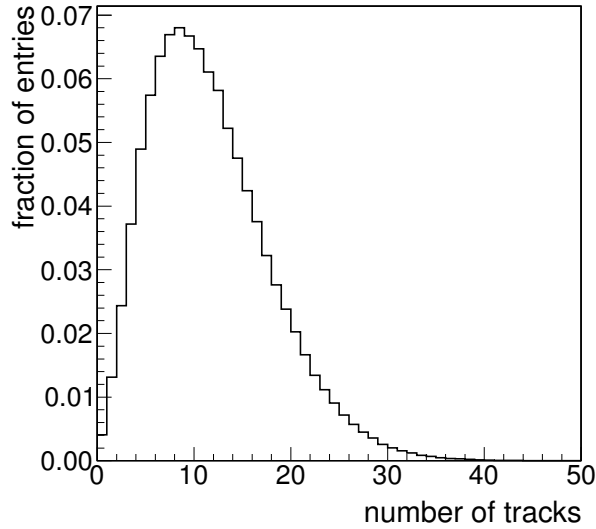
Figure 5.2: Distribution of the number of selected tracks per event on Monte Carlo.

- In the $r - \varphi$ plane, the distance d_0 between the track and the primary vertex (*impact parameter*) is smaller than 5 mm.

- At the point of minimum distance between the track and the origin of the coordinate system, the distance Δz_0 between z_0 of the current track and the average z_0 of all tracks belonging to the same side B_s^0 meson is less than 5 mm: $\Delta z_0 < 5$ mm

The cuts on d_0 and Δz_0 are there to exclude tracks from other interactions within the bunch crossing. The distribution of the number of selected tracks per event can be seen in figure 5.2.

5.3.3 Jet Reconstruction

It is valuable information for flavour tagging to know in which jet a track was reconstructed. For example, the leading fragmentation partner of the same side B meson is most likely found within the same jet.

In top or Higgs physics, jets usually have high transverse momenta p_T and are reconstructed primarily in the calorimeters. In contrast, $b\bar{b}$ jets at the Tevatron are low energetic and are mainly reconstructed in the tracking system. Track based jet reconstruction is done by the *cone clustering algorithm*, using the distance ΔR in $\eta - \varphi$: $\Delta R = \sqrt{\Delta \eta^2 + \Delta \varphi^2}$. Jets are formed by clustering tracks within a cone of $\Delta R < 0.9$, starting at a track with $p_T > 1$ GeV/c .

Figure 5.3: Schematic overview of the opposite side vertexing procedure.

5.3.4 Opposite Side Vertexing

As written above, it is generally not feasible to make an exclusive reconstruction of the opposite side B hadron. However, some information can be gained from inclusive properties of the event. Since B hadrons decays are mediated by the weak interaction, they have a relatively long lifetime, so their decay vertex is often measurably displaced from the production vertex.
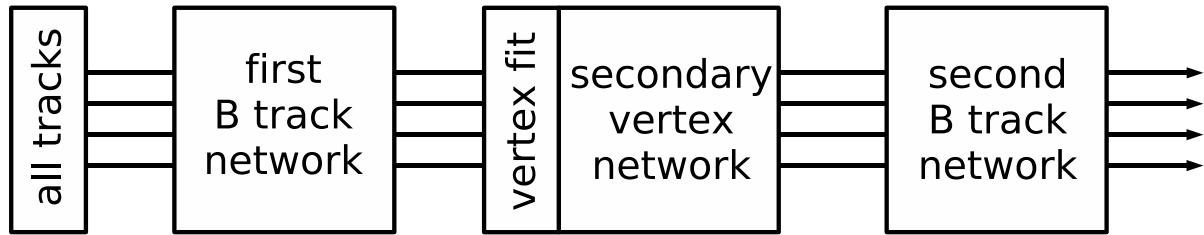
The inclusive reconstruction tries to identify tracks from the B decay chain by means of the *first B track network*. The best track candidates are paired to vertex candidates, which are subsequently ranked in the *secondary vertex network*. With the additional vertex information, the *second B track network* is used to refine the decision of the first stage and to provide a useful track based quantity for later tagging stages. A graphical representation of the procedure can be seen in figure 5.3, the individual steps are described below.

First B Track Network

The purpose of the *first B track network* is to decide for a single track whether it originates from the decay chain of a B hadron or not. Tracks from the B decay chain are regarded as signal in the neural network classification, all others as background. It is trained on and applied to all tracks that fulfil the general track selection criteria listed in subsection 5.3.2. Input variables for this neural network consist of standard track parameters, track parameters with respect to the reconstructed B meson and to the jet as defined in section 5.2.2.

Tracks from secondary vertices have higher values for the impact parameter d_0 and for the longitudinal distance Δz_0 to the primary vertex than other tracks in the event. Standard track parameters are used to benefit from this fact. Also kinematical properties can be exploited: Compared to all background tracks from fragmentation and proton remnants, the B daughter tracks have a relatively high transverse momentum p_T.

In the leading order process *flavour creation*, $b\bar{b}$ quark pairs are created *back to back*, that means their momentum vectors are collinear and of opposite direction. Tracks from the opposite side B decay are therefore expected to accumulate in the opposite spatial
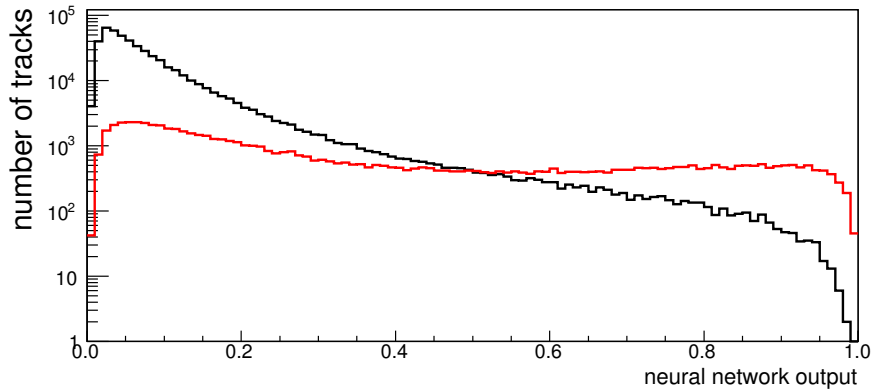
Figure 5.4: Neural network output distribution of the first B track network for tracks from the B decay chain (red) and other tracks (black).

region as the same side B meson. The relation of a given track to the same side B meson is therefore useful information at this stage. Quantities that make this effect visible to the neural network are $\Delta\phi$, ΔR and the signed impact parameter with respect to the same side B meson.

The full list of input variables as well as their importance for the neural network training can be found in appendix A.1. The neural network output distributions are shown in figure 5.4.

Opposite Side Secondary Vertex Fit

All tracks of an event are ranked by their probability to originate from a B decay as given by the first B track network. The best four of them are selected and fitted in pairs to a common vertex, forming six secondary vertex candidates. The invariant mass of the two track vertex is calculated assuming pion mass for both tracks.

The *secondary vertex network* is then used to identify the best B decay vertex candidate. Vertex candidates where both tracks are really from a B decay are treated as signal in the training. Input quantities for this neural network are properties of the two tracks (output from the first B track network, charge, invariant mass), vertex position and fit quality information. The full list is given in appendix A.2.

In the neural network, *bTrackProb2* – the smaller of the B track probabilities from the two tracks – is the most significant input variable, a higher value gives a better vertex probability. It is followed by the decay length significance $L_{xy}/\sigma_{L_{xy}}$ and the *inSameJet* flag. The neural network output distribution is shown in figure 5.5. The vertex candidate with the highest output is selected for the next steps.

In figure 5.6, the performance of the secondary vertexing procedure is evaluated using Monte Carlo truth information. The highest ranked vertex candidate is a true B decay
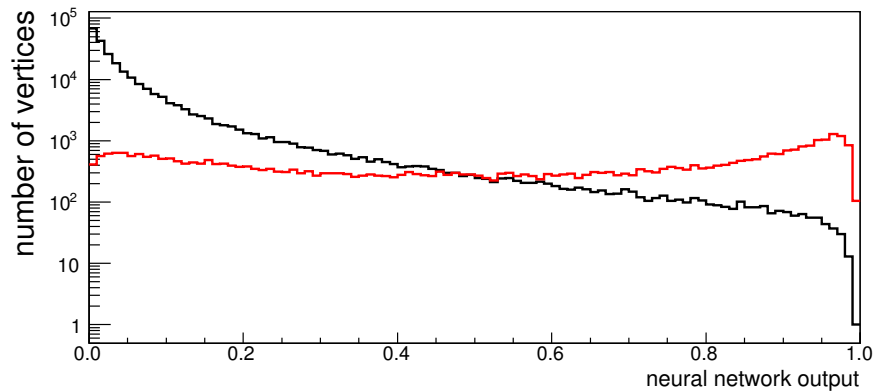
Figure 5.5: Neural network output distribution of the secondary vertex network for true B vertices (red) and wrong candidates (black).
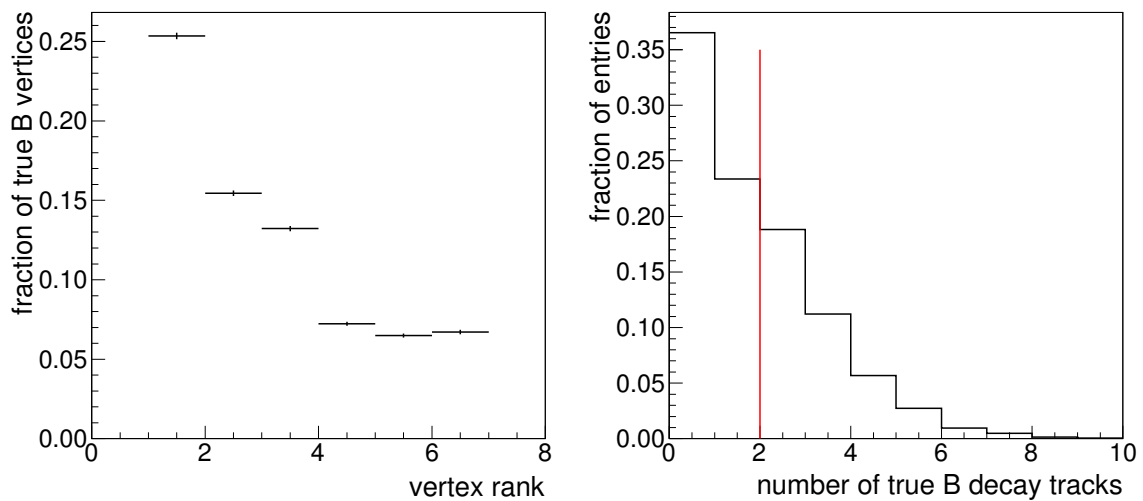
Figure 5.6: Performance evaluation of the opposite side vertexing procedure using Monte Carlo truth information. The left plot shows the fraction of true B decay vertices as a function of the vertex candidate rank given by the secondary vertex network. In the right plot, the distribution of number of true B tracks per event in the tagging track selection is shown. In the region to the left of the red line, less than two true B decay tracks are available, so the reconstruction of a true B vertex is impossible.
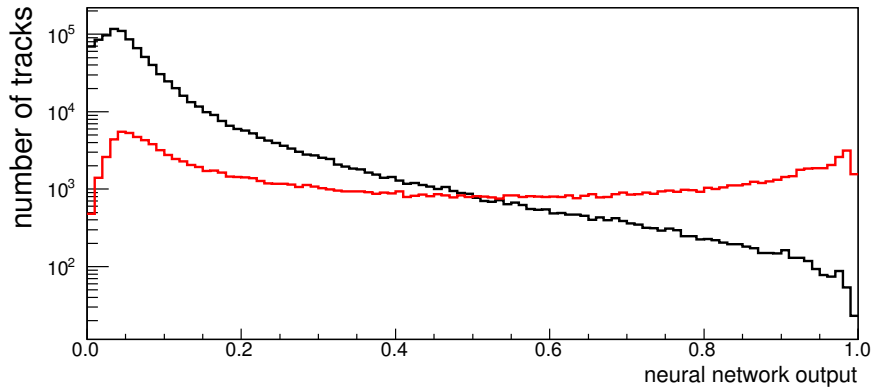
Figure 5.7: Neural network output distribution of the second B track network for tracks from the B decay chain (red) and other tracks (black).

vertex in about 25% of all cases, the following ranks have lower probabilities to be a true B vertex. It can be concluded that the vertexing procedure selects the best available vertex candidate. From the distribution of the number of true B decay tracks in the track selection it becomes evident that it is impossible to find a B vertex in roughly 60% of all cases. A true vertex purity of 25% for the selected candidate where 40% is the maximum achievable value is a very satisfying result.

Second B Track Network

With the information gained about the opposite side decay vertex, another neural network is used to refine the decision of the first B track network. The *second B track network* therefore uses the same signal and background definitions.

It is an obvious choice to use the quantities from the previous vertexing steps as input to this neural network, so the output of the first B track network and the secondary vertex probability are used. Also the location of the secondary vertex should be exploited, so the impact parameter d_0 and the longitudinal distance Δz_0 together with their uncertainties are recalculated with respect to the secondary vertex. A given track's role in the vertex fit is also useful information, so the *usedInFit* and *inJet* flags are used, as well as the track's invariant mass with the vertex, p_{rest} and $\cos\theta_{\mathrm{rest}}$. In addition, muon and electron identification quantities provided by the LeptonTools package are used, and kinematic quantities that were already used in the first B track network like p_T and η are repeated here.

As expected, the output of the first B track network is the most significant input variable, followed by the output of the secondary vertex network and the impact parameter relative to the secondary vertex. The neural network output distributions for signal and background can be seen in figure 5.7, the variable list can be found in appendix A.3.
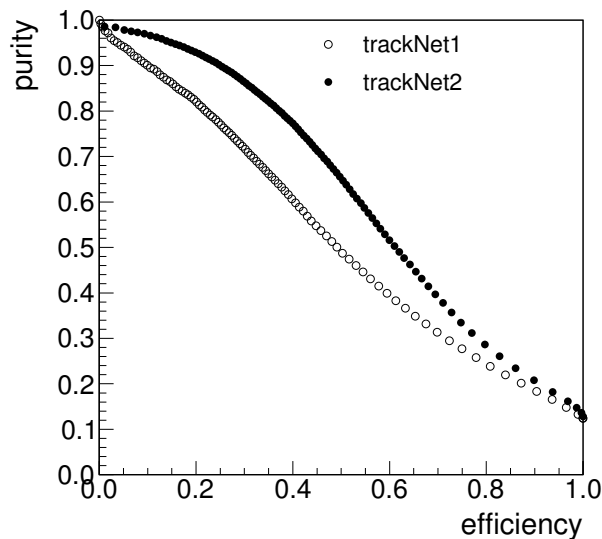
Figure 5.8: Purity as a function of Efficiency for the first (trackNet1) and second (track-Net2) B track network. The definition of purity and efficiency is given in section 4.1.1.

Figure 5.8 illustrates the gain in performance between the first and the second B track network. This neural network output is later used as an input in the track based tagging procedure and as a weight in the vertex charge estimator.

5.3.5 Lepton Identification

Leptons play a prominent role in B physics because semileptonic decays have a noteworthy share of the B branching ratio and leptons provide a good experimental signature. When coming from opposite side B decays, they also carry valuable flavour information. To identify them, the LeptonTools package [46] is used. It provides probability-like variables for the identification of muons and electrons based on artificial neural networks.

Besides interesting electrons from B decays, a lot of electrons originate from photon conversions. In the interaction of photons with matter from the detector, e^+e^- pairs are produced. For most physics analyses, they are background and should be removed. Therefore, the LeptonTools provide an additional neural network for the identification and suppression of conversion electrons.

The idea behind opposite side lepton tagging is described in chapter 2.3.1. To enable the later stages to learn about the valuable lepton tagging information, it is desirable to have a single probability-like quantity indicating whether a given track is a lepton from an opposite side B decay. Providing this is the purpose of the *lepton from B network*.

Leptons from B and from D decays have opposite charge correlations to the flavour

in question, and the ability to distinguish them would further increase the value of lepton flavour tagging. However, both classes have very similar properties compared to background tracks, so only a common identification of both is feasible at this point. Information for a separation between them is included in the track based tagging stages. The signal definition for the lepton from B network contains any true lepton that has a b or $\bar{b}$ quark in its chain of ancestors.

As the signal fraction among all tracks is in the order of 1.5 %, precuts are done to remove most non-lepton background and exclude regions that show suboptimal agreement between data and the simulation (see also section 6.2.1). To be used in the lepton from B network, a track must fulfil one of the following two requirements:

- the muon identification network classifies it as a good muon candidate with $p_\mu > 0.5$

- the candidate has a high output value from the electron identification network and a low conversion probability at the same time: $p_{\mathrm{el}} \cdot (1 - p_{\mathrm{conv}}) > 0.7$

The choice of input variables for this neural network follows the tasks it is designed for: Combining muon and electron identification, suppressing conversion electrons, distinguishing between those from B decays and other leptons. As the second B track network already provides an identification of tracks from B decays, it is used as input here. For lepton identification, the same combinations p_μ and $p_{\mathrm{el}} \cdot (1 - p_{\mathrm{conv}})$ as in the precut are used, both accompanied by a flag if they are defined.

In addition, variables that have already proven useful in the B track networks are used: The relation to the same side B meson is given in form of the signed impact parameter and the relative momenta p_T and p_L.

Opposite side secondary vertex information is also used to find leptons from B decays: Tracks can be transformed into the rest frame of the presumed B decay. Because the two track combination is not a true reconstruction of the full opposite side B decay, its invariant mass is scaled up to 5.3 GeV/c^2 to compensate for the missing components. The following two quantities are calculated in the new vertex rest frame and used in the lepton from B network:

- the track momentum p_{rest} in the secondary vertex rest frame

- the cosine of the helicity angle θ between the track in the vertex rest frame and the corresponding B direction in the laboratory frame $\cos(\theta_{\mathrm{rest}})$

Also the flags *usedInFit* and *inJet* are used as input variables.

The most important input for the neural network is the output of the second B track network, followed by p_T relative to the same side B meson and the signed impact parameter. Lepton identification seems not to play the leading role here because the precut on the lepton identification variables is already quite strict. The neural network output distributions can be seen in figure 5.9, the variable lists can be found in appendix A.4.
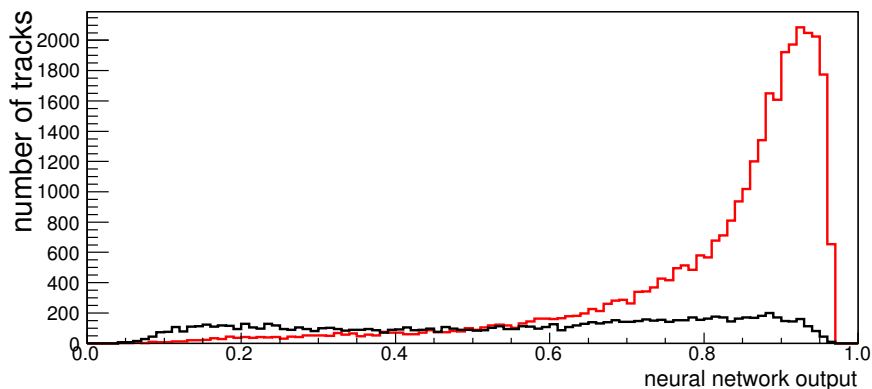
Figure 5.9: Neural network output distribution of the lepton from B network for signal (red) and background (black).

5.4 Track Based Flavour Tagging

The heart of this flavour tagging procedure is to examine all selected tracks for information about the B production flavour. The charge of each track can either have the same or the opposite sign compared to the charge of the b quark in the same side B meson. Track based tagging tries to find correlations and anticorrelations between these charge signs based on a variety of spatial, kinematical and higher level track information.

These correlations are determined in *flavour correlation networks*, artificial neural networks trained to separate correlated from anticorrelated tracks. Signal tracks for these neural networks have the same charge sign as the same side B flavour, background tracks are those where charge and flavour are opposite. The result of the neural network classification is the *track flavour correlation* probability, a continuous value between zero and one. Correlations between track charge and flavour correspond to high values, anticorrelations to low values close to zero. The track flavour correlations are later combined into event level tags.

5.4.1 Subsamples

Following the description in chapter 2.3, flavour information can be gained from several different physical effects. Tracks that are interesting for one effect are often background for another. For example, tracks with high kaon probability are important for same side tagging, but are rather bad lepton candidates. As all sources are small effects by themselves, they need specialized procedures to harvest as much information as possible. Therefore the track charge correlations are not determined in a single neural network. Instead tracks are grouped into three different mutually exclusive *track tagging classes*:

Figure 5.10: Schematic overview of the track based tagging procedure.

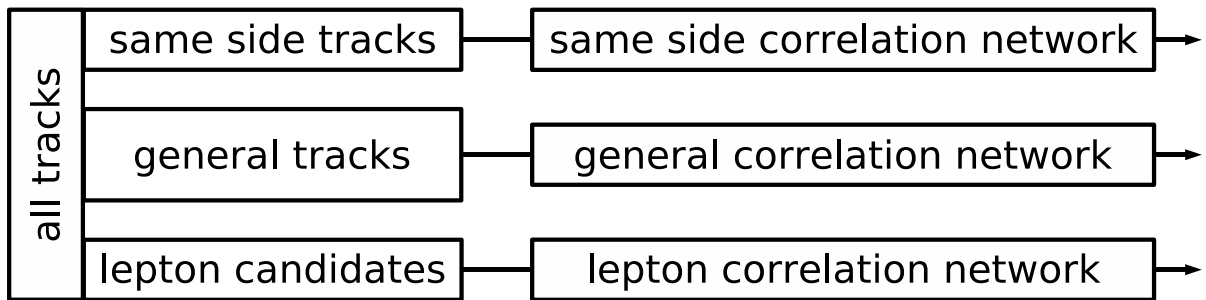- If a track was reconstructed in the same jet as the tracks forming the same side B meson (the *same side jet*), it belongs to the tagging class of *same side tracks* and is handled in the *same side correlation network (SCN)*. The main physical effect that is exploited in this class is the charge of the leading fragmentation particle – a kaon in the B_s case.

- If a track is not part of the same side jet, but passes the precuts for the lepton from B network, it is considered a lepton candidate and handled by the *lepton correlation network (LCN)*. This class is designed to handle opposite side lepton tagging.

- Each track that is not in one of the previous classes is a *general track* and fed into the *general correlation network (GCN)*. This class is not specialised on one tagging effect and tries to catch remaining information like opposite side kaon decays and less pure leptons or same side tracks that do not fit in the relatively tight requirements of the two previous classes.

A schematic overview of the track based tagging step is given in figure 5.10.

Although the three tagging classes have suggestive names and aim at well defined effects, they are not that clean in reality. In the relatively unclean hadronic environment with its many tracks, it is not expected that same side and opposite side can be cleanly separated. Also contributions from the other not aimed-for tagging sources are expected to be present in each of the classes.

5.4.2 Common Set of Input Variables

All track flavour correlation networks are offered the same set of input variables. Their relevance for the individual neural network is determined in the training process. If the contribution of a variable is not larger than 4.5σ, it is completely excluded from the training. The full set of used variables is described here, the corresponding tables for the individual trainings can be found in the appendix.

In the previous stages, high-level quantities have been prepared that already contain distilled information and should be used for the track based tagging step:

- The *B track probability* (bTrackProb2) is the output of the second B track network, which is designed to identify tracks from the opposite side B hadron decay, making use of secondary vertex information. Tracks that are assigned a high B track probability are more likely to carry flavour information than other tracks, which are predominantly background.

- The *lepton from B probability* (lepFromBProb) is the output of the lepton from B network. By its definition, its purpose is to be used for opposite side lepton tagging. As the division into the three tagging classes is done in a hierarchical way, about 10 % of all lepton candidates are contained in the same side tag class, the remaining 90 % form the lepton tag class. Therefore the lepton from B probability is always defined in the lepton tag class and never defined in the general tag class.

To benefit from the reconstructed opposite side B decay vertex, p_{rest} and $\cos(\theta_{\text{rest}})$ are included among the input variables. As most of the vertex information is already contained in the B track probability, these vertex quantities are expected to be most useful in opposite side lepton tagging, where smaller differences after a relatively hard precut on the lepton from B probability can play a role.

Also the relation of a track to the jet in which it was reconstructed can carry information about the flavour. As already exploited in the first B track network, tracks from B decays are usually among the higher momentum tracks forming a jet. The variables included to benefit from such effects are *nTracksJet*, d_0^{jet}, p_T^{rel} and *rapidity* relative to the jet.

The relation to the reconstructed same side B meson is also expected to be useful in finding track flavour correlations. When looking for the fragmentation partner of a meson, it is expected in spatial proximity and will probably have a high momentum and rapidity along the B meson direction. The *signed impact parameter*, $\Delta\eta$, $\Delta\phi$ and ΔR are used to describe the kinematical and and spatial aspects, along with p_L^{rel}, p_L^{rel} *rank*, p_T^{rel} and p_T^{rel} *rank*.

In the following sections, the flavour correlation networks for the three classes of tagging tracks are described. For each of these, physics expectations, plots of important variables and possible interpretations of the training findings are discussed. Details about the conventions and reasons behind the way of plotting these quantities are discussed in chapter 4.5.1. All input quantities as well as their significance for the neural network trainings are given in the appendix.

5.4.3 Same Side Correlation Network

The same side correlation network (SCN) is used for all tracks that are reconstructed in the same side jet. Same side tagging information is predominantly contained in the chain

Figure 5.11: The three most important input quantities for the same side correlation network. The plotting conventions are explained in chapter 4.5.1.

of fragmentation partners of the same side B meson, so information on the kinematical proximity can be expected to play the leading role in this neural network.

As the fragmentation partners for B_s mesons are kaons, also particle identification quantities for pion/kaon separation are expected to be highly useful. However, there are technical and organisatorial issues with these quantities, so the current version of the flavour tagger does not rely on them. This will be discussed further in section 6.5, along with a study on what can be gained.

The three most relevant input variables for the same side correlation network are shown in figure 5.11. Most important is the longitudinal momentum component p_L along the direction of the B meson, showing that higher values correspond to larger flavour correlation probabilities. This is consistent with the expected behaviour for fragmentation tracks of a neutral B meson.

The B track probability turned out to be on second place in importance and shows that probable B daughter tracks tend to be anticorrelated to the same side flavour. Although this tagging class is intended for same side tagging, a clean selection of same side tracks
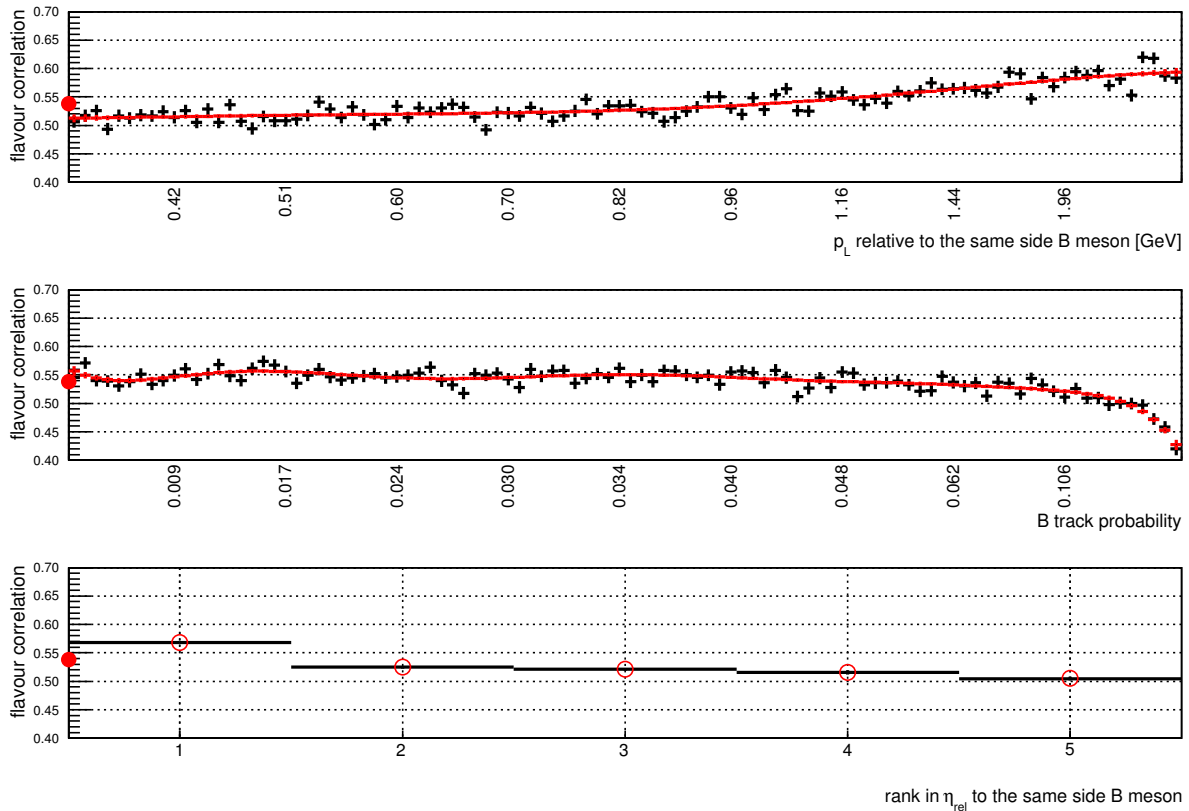
Figure 5.12: The three most important input quantities for the lepton correlation network. The plotting conventions are explained in chapter 4.5.1.

only based on the jet algorithm can not be expected in a hadronic environment. Therefore these tracks most likely originate from the opposite side B decay, for which the observed behaviour is expected.

The third most important variable is the rank in η_{rel} with respect to the same side B meson. The closest track in η_{rel} has a significantly higher flavour correlation than other tracks, which have decreasing correlation probabilities with increasing kinematic distance to the same side B. Again the expectation is confirmed that kinematically close tracks have positive flavour correlations.

The neural network output is part of the combined track flavour correlation probability shown in figure 5.14.

5.4.4 Lepton Correlation Network

Tracks that have a lepton from B probability assigned and do not belong to the same side tagging class form the lepton tagging class and are handled by the lepton correlation
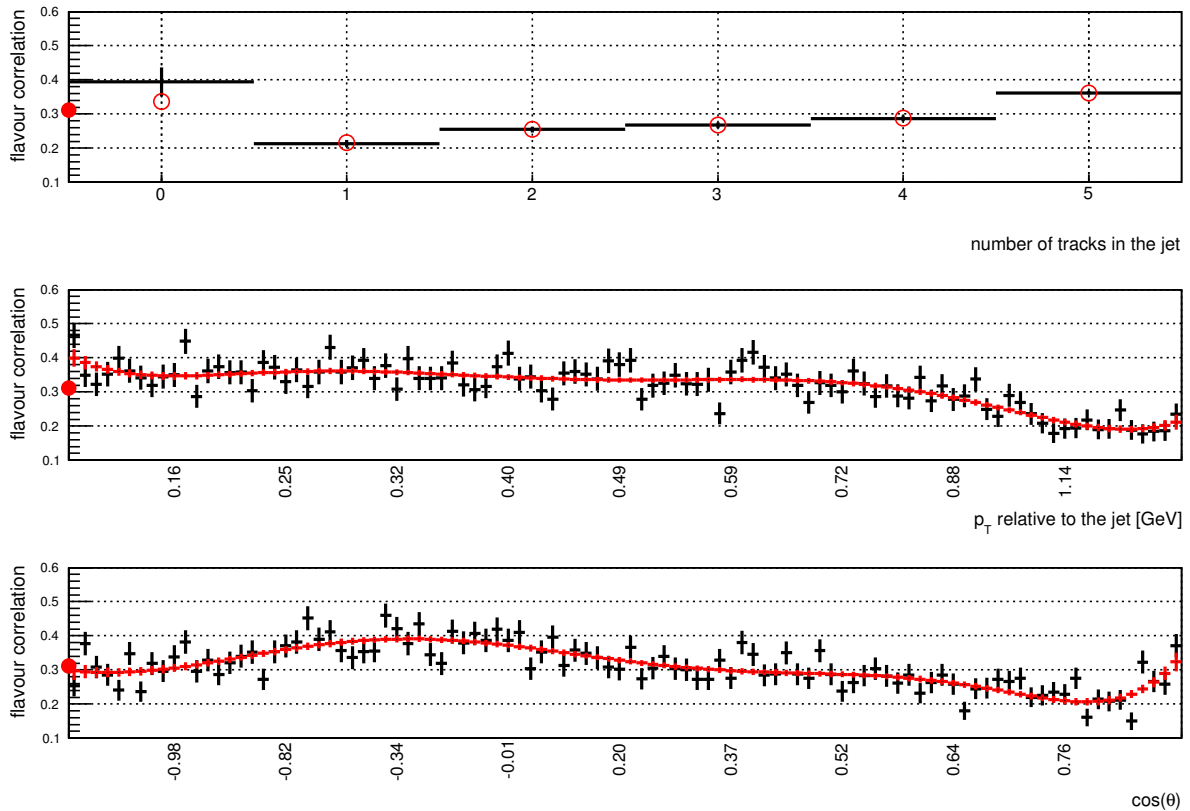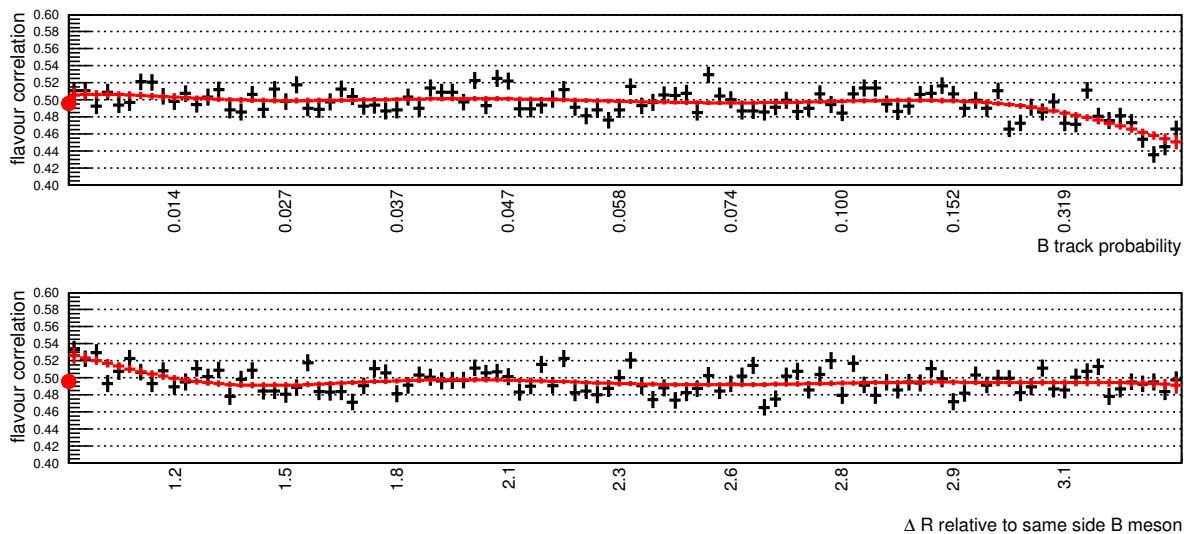
Figure 5.13: Two of the four relevant input quantities for the general correlation network. The plotting conventions are explained in chapter 4.5.1.

network (LCN). Leptons from opposite side B decays have an anticorrelation to the same side flavour. As the prerequirements for the lepton from B network are already quite strict, kinematic quantities to evaluate more subtle effects could play an important role in this neural network.

In figure 5.12, the three most important input quantities for this neural network can be seen. The most important input variable turned out to be the number of tracks in the jet, followed by the transverse momentum relative to the jet and $\cos(\theta)$ between the opposite side vertex momentum direction and the track in the secondary vertex rest frame. Apart from tracks that have not been assigned to a jet (zero tracks in the jet), a lower number of tracks per jet shows a stronger flavour anticorrelation. This might be due to the fact that less misidentified leptons are among the more isolated leptons than among those in the middle of a multi-track jet. A high transverse momentum relative to the jet also indicates a flavour anticorrelation, which is consistent with the expectation for lepton decays.

5.4.5 General Correlation Network

The tracks not fitting in the previous two classes are handled by the general correlation network. The more striking effects from same side and lepton tagging have already been exploited by the other tagging classes, so the power of the general correlation network is expected to be low. Therefore no clear expectations corresponding to special physics effects can be given in advance. The same selection of input quantities as for the other classes is offered in the hope that inclusive effects might be still relevant.
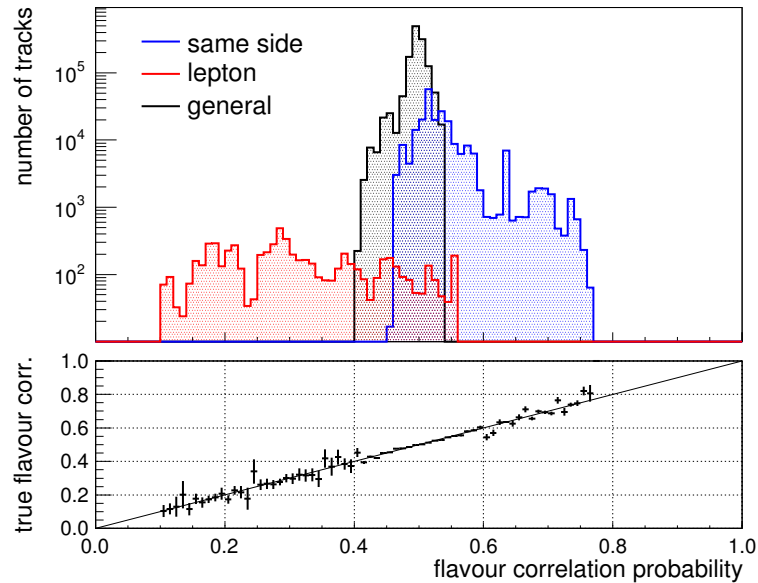
Figure 5.14: Neural network output distributions of the three track flavour correlation networks for the different tagging classes. The distribution in each class is the sum of signal (correlated tracks) and background (anticorrelated tracks) distributions. The lower plot shows that the neural network output can be regarded as a probability value for the true flavour correlation.

Figure 5.13 shows two of the four variables that were found to be relevant enough for the classification process by NeuroBayes preprocessing. The most important quantity is the track from B probability, showing the same behaviour as in the same side correlation network. As there is no special tagging class for non-lepton opposite side B tracks, the importance of this high-level input quantity is not surprising. The other input variable in this figure, ΔR to the same side B meson, is the fourth most important for this neural network. The assignment of tracks to the same side tagging class is based on the findings of the jet clustering algorithm, so it is a reasonable assumption that same side tagging effects can also be found in this class of remaining tracks; to illustrate this, this input was chosen for plotting here. Tracks that are close to the same side B meson are likely to have positive flavour correlations.

In figure 5.14, the neural network output distributions for all three flavour correlation networks is shown. As expected, lepton tracks have show an anticorrelation between charge and flavour, while same side tracks tend to have a positive flavour correlation. The tracks of the general class accumulate in the central region, yielding the lowest separation power of the three classes.
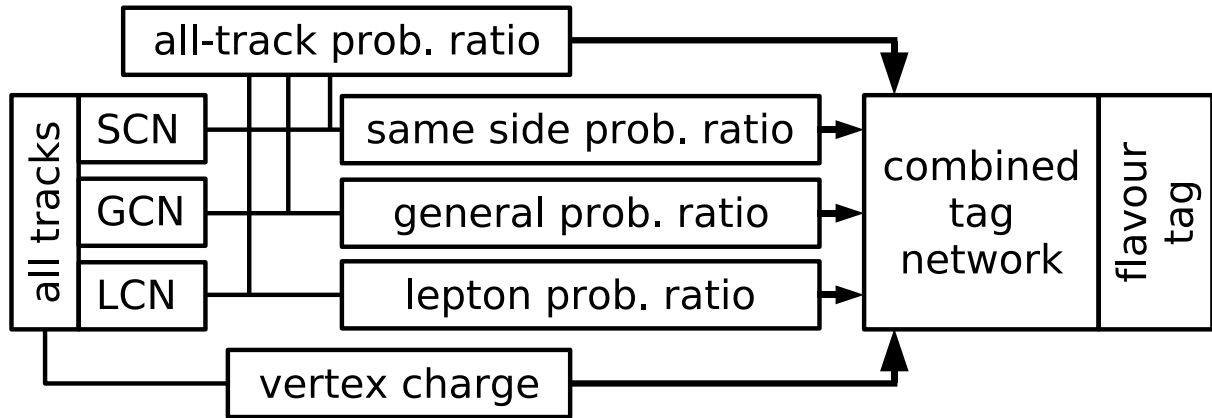
Figure 5.15: Schematic overview of the event level tag combination process.

5.5 Flavour Tagging on Event Level

All tagging efforts described so far are based on single tracks. An event level flavour tag is obtained by combining the information of multiple tracks. This is done with four probability ratios of the track flavour correlation probabilities and an estimator of the opposite side vertex charge. All event level tags are combined in a neural network, delivering the final decision of the flavour tagger containing all information available. A graphical overview over the event level tag combination is given in figure 5.15.

5.5.1 Probability Ratios

In first order, all track flavour correlation probabilities can be regarded as independent measurements which can be combined in a probability ratio. For each track i, the flavour correlation probability p_i is a prediction whether the track charge q_i has the same sign as the B meson flavour to tag. To transform this relative flavour correlation into an absolute statement on the flavour, the track charge needs to be taken into account. There are two cases for the probability p_i^B to have a B meson, containing a $\bar{b}$ quark:

$$p_i^B = p_i \quad \text{for} \quad q_i = -1 \tag{5.1}$$

$$p_i^B = 1 - p_i \quad \text{for} \quad q_i = +1 \tag{5.2}$$

The probability P_B on event level to have a B meson is obtained by multiplying track probabilities and normalizing the overall probability to 1:

$$P_B = \frac{\prod_i p_i^B}{\prod_i p_i^B + \prod_i (1 - p_i^B)} \tag{5.3}$$
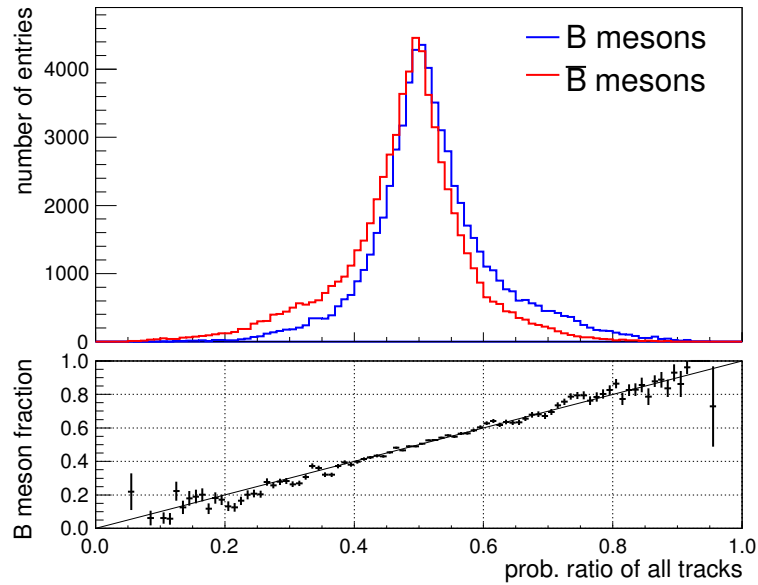
Figure 5.16: Distribution of the probability ratio of track flavour correlations using all tracks.

A low probability to have a B meson is equivalent to a high probability to have a $\bar{B}$ meson:

$$P_{\bar{B}} = 1 - P_B \tag{5.4}$$

Besides a probability ratio using all selected tracks in an event, also separate ones are calculated in each of the three tagging classes. Although it can be expected that the combination of all tracks contains all available flavour information, the additional probability ratios give insight into the relative contributions of the tagging classes to the overall power. Furthermore, each event level tagging quantity can in principle be used for verification of the procedure and of agreement between simulation and reality. Validation of the individual event level tags on Monte Carlo is further discussed in section 6.1.2. The distribution of the track probability ratio using all tracks can be seen in figure 5.16.

5.5.2 Opposite Side Vertex Charge

Another quantity for flavour tagging on event level can be calculated from the B track probability. By calculating a weighted sum of the charges q_i of all tracks i in an event, using the B track probability p_i^B as weight, an estimator Q_V for the charge of the reconstructed opposite side B decay vertex is obtained:

$$Q_V = \sum_i p_i^B \cdot q_i \tag{5.5}$$
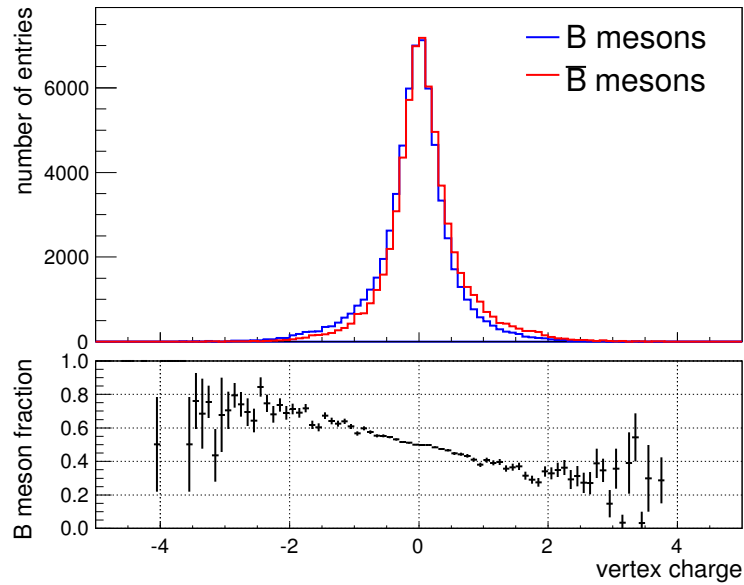
Figure 5.17: Distribution of the vertex charge estimator and correlation to the same side B flavour.

The number of tracks per event differs, its distribution is shown in figure 5.2. A normalization of the vertex charge to the number of tracks involved is not done. The Q_V distribution and its correlation to the same side B flavour can be seen in figure 5.17. A low value for Q_V corresponds to a high probability of having a B meson on the same side.

5.5.3 Combining All Tags

In the last step of the tagging procedure, a final neural network is used to combine the four track probability ratios with the vertex charge estimator. The output of this neural network is a probability value that the same side B meson is a particle containing a $\bar{b}$ quark, as opposed to being an antiparticle containing a b quark, and can be regarded as the result of the whole flavour tagging procedure. Its distribution can be seen in figure 5.18.

Performance measurements on simulated events, validation, cross checks with B^0 and B^+ mesons and results on data are presented and discussed in the next chapter.
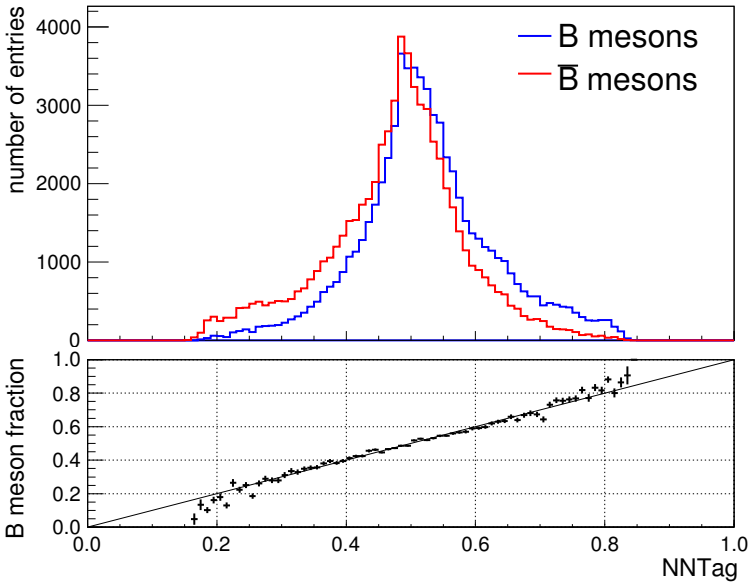
Figure 5.18: Output distribution of the final neural network (NNTag) used to combine all available tagging information on event level.

6 Validation

On the way to a reliable flavour tagger for application in B physics analyses, the development result described in the previous chapter is an important milestone. However, Monte Carlo simulations as those used for this tagger always have a limited degree of realism. A part of this is because the complicated simulation of all detector aspects must be achieved within reasonable development and computing time, and also because we are just at the borders of knowledge about the physics processes that are involved. Therefore Monte Carlo predictions should be compared to findings on data wherever possible to find out how far the procedure can be trusted and to quantify its reliability.

In this chapter, the different items of the validation procedure and their results are described. The first step is to evaluate the training results on the B_s^0 Monte Carlo and determine the predicted tagging power for the subtaggers and the neural network combination. Then the agreement between the Monte Carlo simulation and real data is examined and measures are described that were taken to improve the agreement where necessary. After that, cross checks for the procedure on B^+ and B^0 mesons and their results are provided. At the end, preliminary results of the new B_s^0 mixing analysis are discussed and an outlook on possible improvements to this flavour tagger by particle identification information is given.

6.1 Performance on B_s^0 Monte Carlo

The first step towards validation of the tagger is to evaluate its performance on a Monte Carlo simulation similar to the one used for training. A sample independent of the actual training sample, the *validation Monte Carlo sample*, was used where possible. If there are overtraining effects, they will become evident within the validation procedure. Only the lepton part is an exception to this: All available lepton candidate tracks on Monte Carlo have been used for training the *lepton from B* and the *lepton correlation networks*, so unfortunately no independent sample is left to test this part on Monte Carlo.

6.1.1 Combined Neural Network Tagger

The final output of the tagger is a continuous value between zero and one that can be interpreted as a probability that the same side B is a particle, containing a $\bar{b}$ quark. To verify the behaviour of this output it is useful to plot the distributions of this quantity
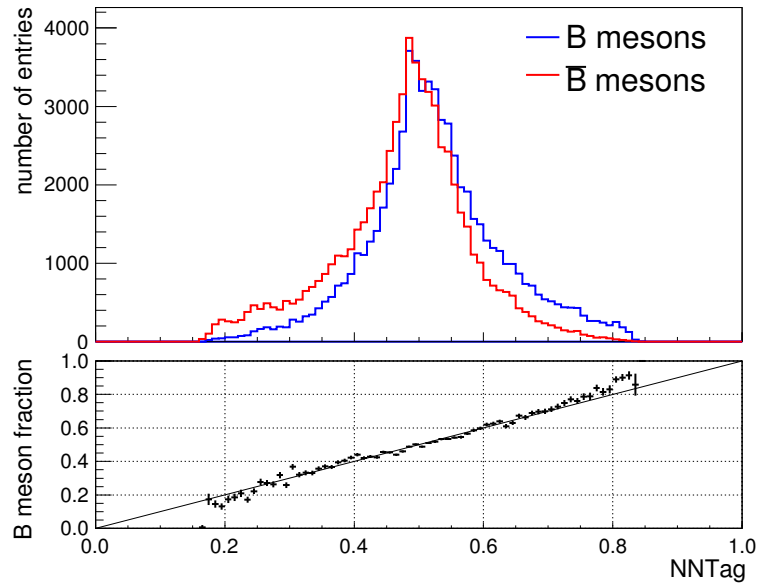
Figure 6.1: Output distribution of the combined Neural Network tagger for true B and true $\bar{B}$ mesons on the validation Monte Carlo sample.

for true B mesons and for true $\bar{B}$ mesons separately. Particles should accumulate at higher, antiparticles at lower output values. Furthermore, the probability behaviour can be checked by plotting the fraction of true B mesons in every bin. For a good estimator all points in this plot should be on a straight line with slope one. The described plots can be seen in figure 6.1.

The output distribution has a clear peaking shape in the central region between 0.4 and 0.6, where events accumulate that can not be tagged very clearly. The lower tail extends to values around 0.15, the higher tail to about 0.85. These regions are populated by events where the best decisions can be made.

The lower plot shows the relation between the neural network output and the fraction of B mesons in a given bin. The plotted line represents the expectation for an ideally trained neural network. Small systematic deviations can be seen in the tail regions, so the classification has slightly more separation power there than the actual output value suggests. The overall agreement between the expected and measured behaviour is very good, especially in the central region of the neural network output, where most events are found.

The distributions are not completely symmetric to each other, which means that both production flavours can be identified with different uncertainties. There are physical as well as detector effects that are different for tracks of the two charges: The cross sections for positive and negative kaons for interaction with material are different, and the COT geometry leads to different tracking efficiencies for positive and negative tracks.

Charge symmetry is not assumed or enforced in the tagging procedure, so the resulting asymmetries give no reason to worry.

The definition of the dilution from equation 2.29 can be written using the probability P_{RS} to have a right tag:

$$D = \frac{N_{RS} - N_{WS}}{N_{RS} + N_{WS}} = \frac{2 \cdot N_{RS}}{N_{RS} + N_{WS}} - 1 = 2 \cdot P_{RS} - 1. \tag{6.1}$$

The neural network output is a probability P_B to have a B meson. Values below 0.5 are considered $\bar{B}$ tags, the others B tags. The probability P_{RS} to take the right decision is therefore equal to P_B in case of a B meson and $1 - P_B$ in case of a $\bar{B}$ meson. The tagging power $\mathcal{T}$ is evaluated using the event by event predicted dilution:

$$\mathcal{T} = \frac{1}{N} \sum_{i=1}^{N} D_i^2, \tag{6.2}$$

where N is the total number of events and D_i the predicted dilution of the individual event. On the validation Monte Carlo sample, the tagging power was measured to be

$$\epsilon = 99.6\% \tag{6.3}$$
$$D^2 = 4.6\% \tag{6.4}$$
$$\mathcal{T} = \epsilon D^2 = 4.6\%. \tag{6.5}$$

The combined neural network tagger provides a tagging decision for every event that has at least one track not used in the same side B reconstruction, so its overall efficiency is close to one. For the subtaggers that do not have this feature, events without a tagging decision are equivalent to $D_i = 0$.

6.1.2 Individual Track Probability Ratios

It is worthwhile to check not only the final combination of all information, but as well the contributing input quantities on event level, as it gives insight into their share of the overall tagging power and can highlight possible problems in single parts. As described in section 5.5.1, the track flavour correlation probabilities are combined in three separate probability ratios for the three track tagging classes as well as in one probability ratio for all available tracks. The behaviour of these probability ratios on the validation Monte Carlo sample is examined here. Their output distributions can be seen in figure 6.2.

The same side tagging track class has the highest tagging power of the three classes. In the majority of events, a tag in this class is available, and the same side track kinematics exploited in the corresponding flavour correlation network yield reasonable discrimination
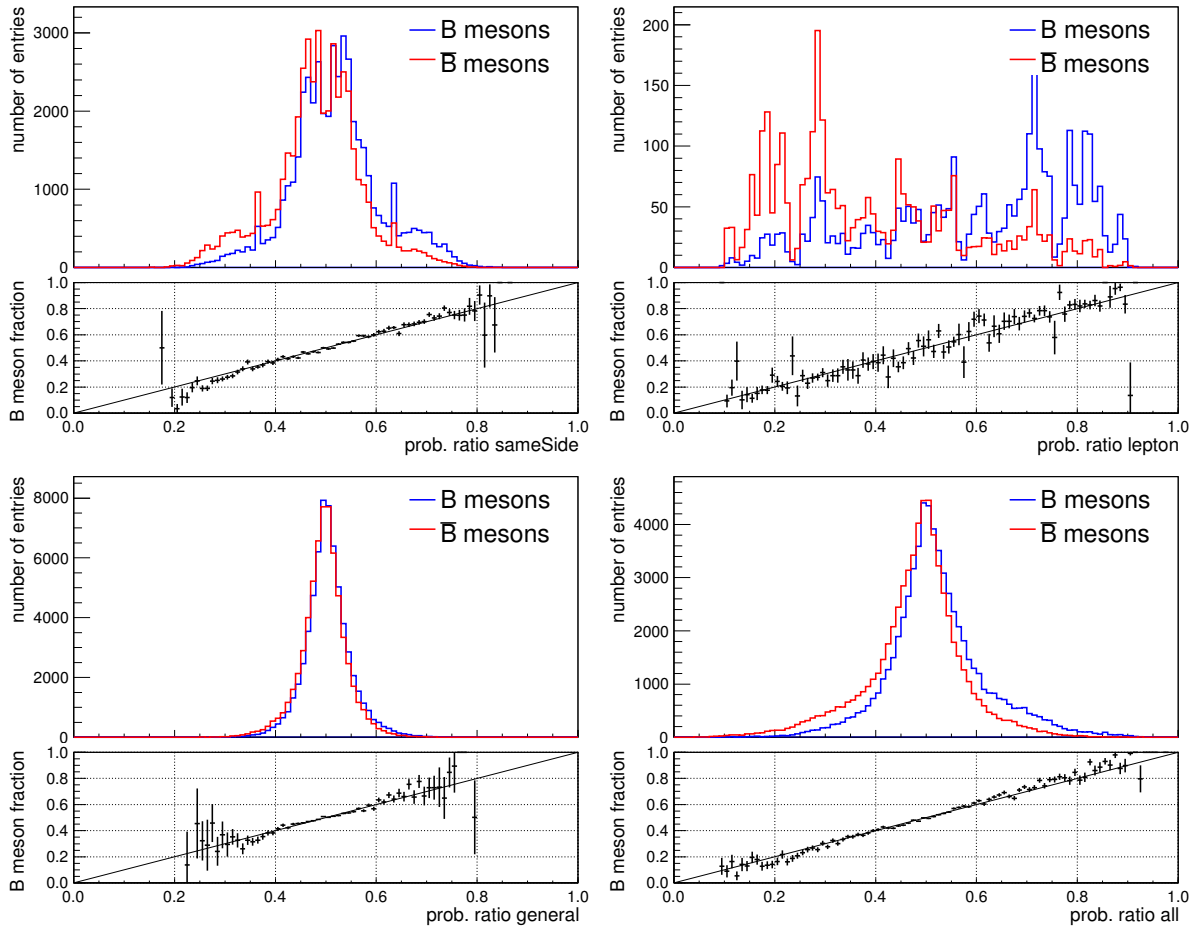
Figure 6.2: Distributions of the individual track probability ratios for true B and true $\bar{B}$ mesons on the validation Monte Carlo sample. In the upper line, the probability ratios for the same side (left) and lepton (right) classes can be seen, below those for the general track class (left) and for all tracks (right) are shown.

power between both flavours. Its tagging power was measured to be

$$\epsilon = 71.5\% \tag{6.6}$$
$$D^2 = 3.3\% \tag{6.7}$$
$$\mathcal{T} = \epsilon D^2 = 2.4\%. \tag{6.8}$$

In the lepton tagging class, quite clean decisions can be made, but only a small fraction of all events have at least one lepton candidate track:

$$\epsilon = 4.7\% \tag{6.9}$$
$$D^2 = 19.6\% \tag{6.10}$$
$$\mathcal{T} = \epsilon D^2 = 0.9\%. \tag{6.11}$$

This behaviour is amplified by the strong requirements that had to be done for the lepton tagging class due to suboptimal agreement between data and Monte Carlo (see section 6.2.1).

In nearly all events there are tracks that belong to the general tagging class. Because the well measurable effects have been treated by the other two tagging classes, the power is relatively low. It was measured to be

$$\epsilon = 99.2\% \tag{6.12}$$
$$D^2 = 0.8\% \tag{6.13}$$
$$\mathcal{T} = \epsilon D^2 = 0.8\%. \tag{6.14}$$

The probability ratio of all tracks contains all information of the three subclasses and is defined for each event with at least one track not used in the same side B reconstruction. Its tagging power

$$\epsilon = 99.6\% \tag{6.15}$$
$$D^2 = 4.0\% \tag{6.16}$$
$$\mathcal{T} = \epsilon D^2 = 4.0\% \tag{6.17}$$

corresponds to the sum of the values in the individual classes.

Also the relation between the B meson probability predicted by the respective tagger and the true B meson fraction can be seen in figure 6.2. For all four cases the observed behaviour is consistent with the ideal behaviour indicated by the diagonal line in the plot. No parametrization as for the vertex charge was needed to achieve this feature. The assumption that the individual track flavour correlations are uncorrelated measurements can be regarded as justified. Therefore the results of the track probability ratios can be interpreted as B meson probabilities.
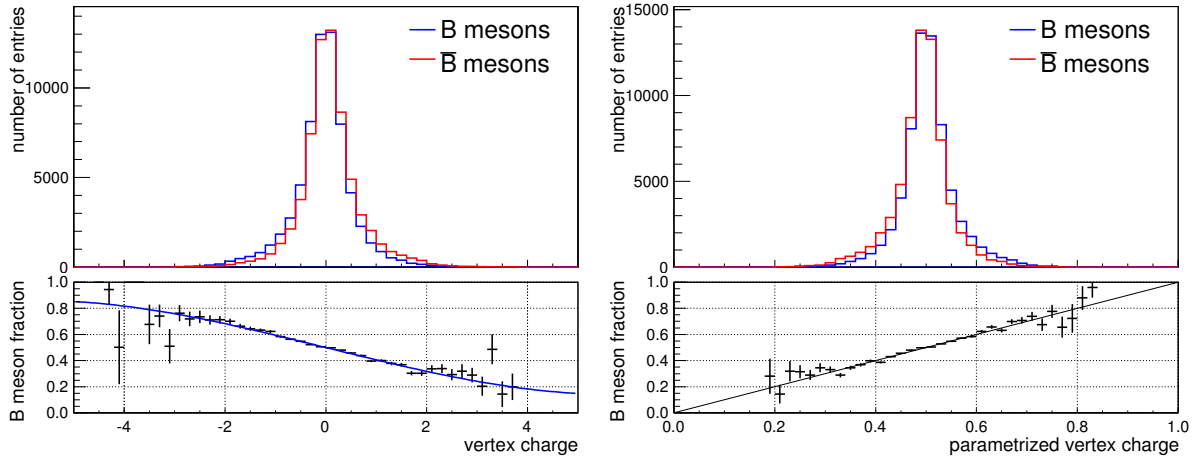
Figure 6.3: Distribution of the vertex charge tag in its raw form (left) and after the parametrization fit (right). The fit function is shown in the lower left plot.

6.1.3 Vertex Charge Estimator

In contrast to the other event level tagging quantities, the vertex charge estimator is not based on a track probability ratio combining the output of flavour correlation networks. As explained in section 5.5.2, it is a weighted sum of a varying number of track charges without a special normalization. To transform it into a probability for having a B meson similar to the other event level tags, a parametrization fit is needed. This is only necessary when using the vertex charge as a stand alone flavour tagger, for example for checking its performance. When it is used as an input for the final tagging neural network, the raw form is fully sufficient because of NeuroBayes preprocessing.

The raw vertex charge (shown in figure 5.17) distribution is centered at zero and extends from roughly -4 to 4. As parametrization fit function,

$$f(x) = 0.5 + a \cdot x + b \cdot x^3 \tag{6.18}$$

is chosen, where x corresponds to the raw vertex charge value and $f(x)$ to the parametrized vertex charge. The mapping of the neutral point from 0 in the raw form to 0.5 in the probability form is fixed to avoid random shifts. To ensure that no asymmetries are introduced by the fit, the function is symmetric, leaving out a possible quadratic term. The cubic term is meant to catch outliers that might fake extraordinary high dilutions when only mapping with a linear relation. In figure 6.3, the distribution of the raw vertex charge, the parametrization fit and behaviour of the transformed quantity are shown. It can be seen that the cubic term in the fit function is also justified by the shape of the B

meson fraction. The parameters are determined in the fit to be

$$a = (-9.5 \pm 0.3) \cdot 10^{-2}, \tag{6.19}$$
$$b = (9.9 \pm 7.3) \cdot 10^{-4}. \tag{6.20}$$

After the transformation, also the vertex charge can be regarded as a B meson probability, and the tagging power was determined to be

$$\epsilon = 99.6\% \tag{6.21}$$
$$D^2 = 1.3\% \tag{6.22}$$
$$\mathcal{T} = \epsilon D^2 = 1.3\% \tag{6.23}$$

Also the vertex charge is available for every event with at least one track not used in the same side B reconstruction, leading to 99.6% efficiency.

6.2 Agreement between Data and Simulation

6.2.1 Lepton Identification

During development of the flavour tagger, notable differences between the Monte Carlo simulation and real data have been observed for the lepton identification quantities. The LeptonTools have been developed and tested several years ago. Since then, the trigger system and the offline reconstruction software have evolved. The work needed to investigate and improve the observed disagreement in detail is beyond reach as a part of tagger development. Fortunately the discrepancies predominantly affect the regions of low signal fraction, so the problem can be attenuated by tightening the selection that needs to be done anyway.

The comparison plots of electron and muon identification quantities can be seen in figure 6.4. Especially for $p_{\mathrm{muon}} < 0.5$, a clear discrepancy between data and simulation can be seen. The region above 0.5 shows roughly the same shape, so only muon candidates with $p_{\mathrm{muon}} > 0.5$ are used.

In terms of statistics, the lepton from B network and the lepton correlation network are at the lower limit and would benefit from softer precuts. For future updates of the flavour tagger it would be desirable to either have updated LeptonTools available or switch to the likelihood based electron and muon identification used in the existing opposite side lepton taggers.
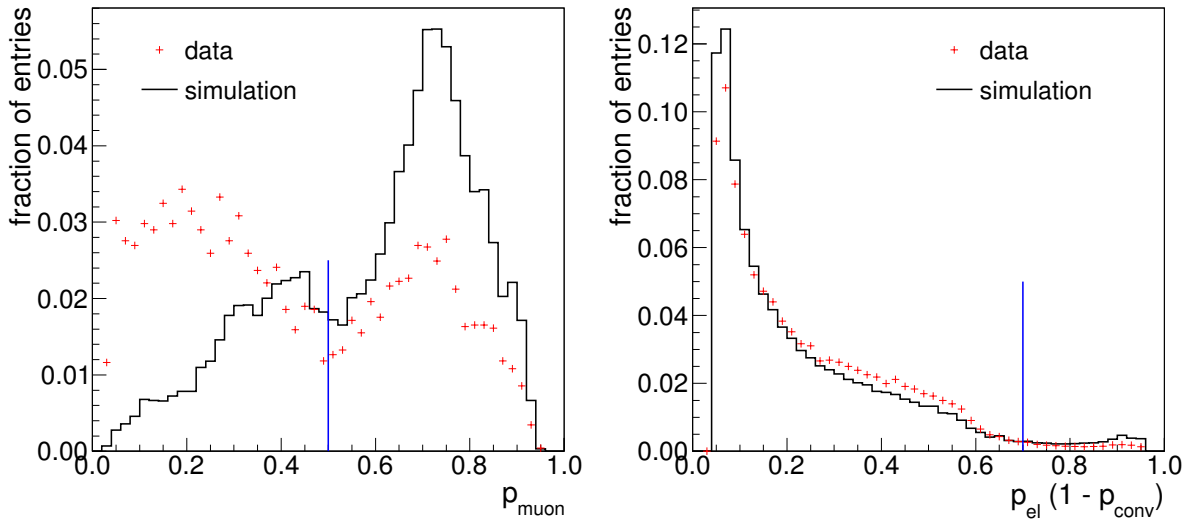
Figure 6.4: Data to Monte Carlo comparison for the lepton identification quantities provided by the LeptonTools package, shown on B_s^0 data and Monte Carlo. The muon probability p_{muon} is shown on the left, the combination of electron probability p_{el} and conversion probability p_{conv} that is used in the tagger can be seen on the right. The precuts chosen to avoid the problematic regions are indicated with the vertical lines.

6.2.2 Multiple Interactions per Bunch Crossing

Since the beginning of Run II, the Tevatron has significantly increased its instantaneous luminosity. At the time when the CDF software was designed, it was a reasonable assumption that there is only one hard scattering process per bunch crossing, so the standard Monte Carlo simulations do not have multiple interactions. A so called *pile-up* of multiple interactions is to be expected in today's luminosity regime. Tracks from these additional interactions that pass the track selection criteria for this flavour tagger are completely unrelated to the B meson of interest. They can however fake signatures of interesting tracks, for example appear like coming from a secondary vertex due to a large impact parameter.

The purpose of this study is to quantify the amount of pile-up tracks that are to be expected in order to decide on additional measures against these tracks if necessary. It was performed on the B^0 version of the flavour tagger, but is equally valid for the B_s^0 and B^+ variants. The number of hard scattering processes per bunch crossing is calculated from Tevatron parameters and the interaction cross section. From the Tevatron radius of 1km, the revolution time is calculated to be

$$T = 1 \text{ km} \cdot \frac{2\pi}{\mathrm{c}} \approx 21 \ \mu s. \tag{6.24}$$
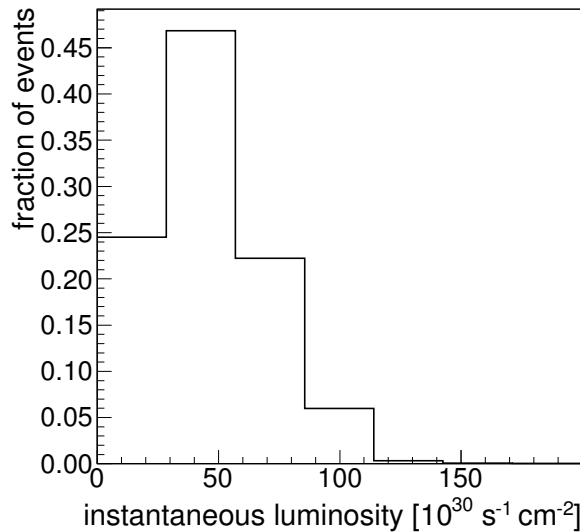
Figure 6.5: The distribution of instantaneous luminosity at which the events in the B^0 data sample have been recorded. The bin width corresponds to $\Delta\mathcal{L} = 28.5 \cdot 10^{30}$ sec^{-1} cm^{-2}, the value for which the average number of expected primary interactions increases by one.

CDF has measured the total inelastic $p\bar{p}$ cross section to be $\sigma = 60.3$ mb [52]. Taking into account that the Tevatron is filled with 36 pairs of colliding bunches, the number of interactions n per bunch crossing as a function of instantaneous luminosity $\mathcal{L}$ is determined by the relation for the event rate

$$\dot{N} = \frac{n \cdot 36}{T} = \mathcal{L} \cdot \sigma, \tag{6.25}$$

which means that one additional interaction is expected for an increase of

$$\Delta\mathcal{L} = 28.5 \cdot 10^{30} \text{ sec}^{-1} \text{ cm}^{-2} \tag{6.26}$$

in instantaneous luminosity.

According to the track selection criteria listed in section 5.3.2, tracks are selected for tagging when they are within tight d_0 and Δz_0 limits around the reconstructed same side B meson. A thorough way of determining how many tracks from additional interactions actually fulfil these requirements is to make a complete Monte Carlo simulation including additional interactions. This is done by using the same simulation process as described in section 5.3.1, but inserting a specified number of minimum bias interactions generated by Pythia [50] before detector simulation. Our dataset of reconstructed B^0 mesons was used to extract the distribution of instantaneous luminosity in bins with a width of $\Delta\mathcal{L}$
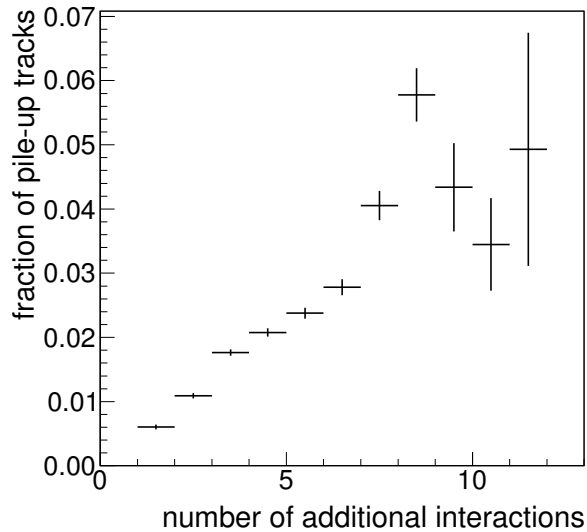
Figure 6.6: Result of the pile-up Monte Carlo study: The fraction of pile-up tracks among the selected tracks in an event as a function of the number of additional interactions that were simulated.

calculated above. The luminosity profile in this binning can be seen in figure 6.5.

Another Monte Carlo sample including the simulation of pile-up interactions is then created with this luminosity profile. How many additional interactions are generated for a given event is determined by a random number that follows a poisson distribution with the mean of the given luminosity region. Effects of pile-up corresponding to the behaviour in data are included in this enhanced Monte Carlo sample.

Using Monte Carlo truth information, the fraction of pile-up track passing the selection criteria can be determined as a function of the number of additional interactions simulated. This relation is shown in figure 6.6. While the fraction of pile-up tracks is increasing with additional interactions, it is still below 3% for nearly all simulated events. Taking into account the distribution of selected tracks per events (see figure 5.2), one additional pile-up track is expected on average in every third event. It is unlikely that pile-up effects at such low intensities affect the performance of the flavour tagger notably. Therefore they are regarded to be negligible, and no special Monte Carlo simulations for B_s^0 and B^+ have been generated. As the B^0 Monte Carlo sample has been created within this study, it is used for training of the B^0 version of the flavour tagger.

6.2.3 Reweighting

Some of the inherent limitations of Monte Carlo simulations can be overcome by weighting the individual Monte Carlo events in a suitable way to achieve the desired properties. The

correction weights can have continuous values, but their distribution should not extend across several orders of magnitude. It is also reassuring if the average weight is around one, because then the sum of all weights equals the number of events before reweighting. In the following sections, reweighting measures for improving the Monte Carlo used in the flavour tagger are described.

Opposite Side B Hadron Composition

In the b quark fragmentation process, about 40% B^0 , 40% B^+ , 10% B_s^0 and 10% other B hadrons are produced [14]. When generating the Monte Carlo according to the procedure described in section 5.3.1, this composition is distorted and needs to be restored by appropriate weights. The generation process involves filtering of events containing at least one of the desired B meson type, which is assumed to be a B_s^0 meson in the context of this description, but is valid in the same way for the other two B meson types. Exactly one B_s^0 meson per event is then forced to decay in the selected decay channel, artificially enhancing the branching ratio of this channel to one. The selected decay channel in the B_s^0 case is

$$B_s^0 \to D_s^\pm \pi^\mp, \ D_s^\pm \to \phi \pi^\pm, \ \phi \to K^+ K^- \tag{6.27}$$

with a branching ratio $\mathcal{B} = 7.2 \cdot 10^{-5}$ [14], including the subsequent $D_s^\pm$ and ϕ decays. The artificial branching fraction enhancement saves a considerable amount of computing time.

If there is a second B_s^0 meson present, its decay is simulated according to the decay table without additional constraints. Which of the two B_s^0 mesons decays freely is decided randomly, with equal probabilities for both. In comparison to the case of only one B_s^0 meson per event, the branching ratio of the selected decay channel is effectively enhanced to $\frac{1}{2}$ instead of one.

After Monte Carlo generation, B_s^0 mesons in the specified decay channel are reconstructed, defining the *same side B meson*. Now first consider all cases where the $\bar{b}$ quark enters the same side B: All events with one B_s^0 meson are reconstructed, but only half of those with two B_s^0 mesons[1]. The other case with the b quark on the same side is symmetric, again all events with one B_s^0 meson are found, but only half of those with two B_s^0 mesons. For the opposite side B hadron, only half of the possible cases for B_s^0 are retained, while all other pass unrestricted.

To correct this distortion, all events that have the same kind of B meson on the same side and on the opposite side are weighted up by a factor of two in relation to all other events. The sum of all weights is normalized to the number of events in order to keep the total number of events unchanged. The composition of opposite side B hadrons before and after reweighting is shown in figure 6.7 for the B^0 case.

---

[1] To be precise, also the second B can decay in the preferred channel, so it would be $0.5 + \mathcal{B}$, which is a negligible correction in this case.
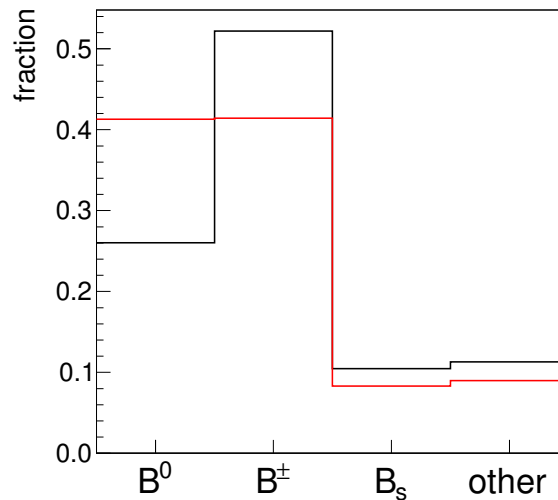
Figure 6.7: Composition of opposite side B hadrons before (black) and after (red) the described reweighting procedure, shown for the B^0 case.

Trigger

The CDF trigger system is continuously evolving to accommodate external influence like the increased Tevatron luminosity as well as to benefit from grown understanding of the system and its suitability for analysis needs. The overall bandwidth is limited by the data acquisition system, so all trigger scenarios are competing for their share of the total bandwidth. The balance between them is maintained by *prescaling factors* which are adjusted in a global optimization of technical and physical requirements. Luminosity dependent prescales are used to enable high-rate triggers only after the instantaneous luminosity has dropped below specified threshold values.

As the whole trigger system is very complex, it is very difficult to predict the available free bandwidth in a given operational situation. Offline determination of static prescaling factors can therefore only be used as a starting point for further optimization. To make optimal use of the available bandwidth, a *dynamic prescaling system* is used to grant individual triggers more trigger rate if free bandwidth is available.

No simulation of the dynamic prescaling system is available for Monte Carlo production. The composition of trigger scenarios in the Monte Carlo dataset can therefore be significantly different from the one observed in real data. As the trigger scenario in which an event was recorded influences quantities like the momentum spectrum of reconstructed B mesons, data to Monte Carlo disagreement in trigger composition should be corrected before examining the agreement in kinematic properties.

In figure 6.8, the distribution of trigger scenarios on data and on Monte Carlo before
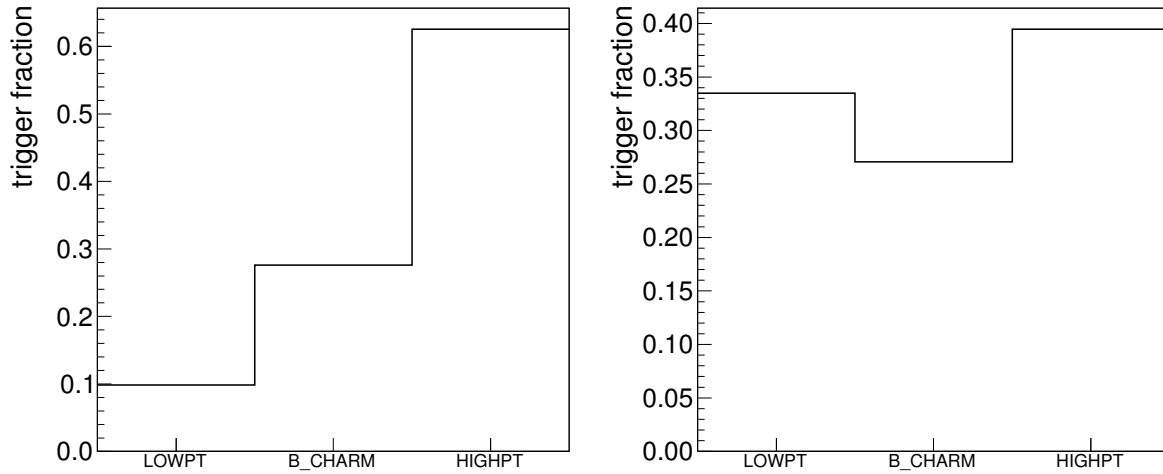
Figure 6.8: Composition of trigger scenarios for B_s^0 data (left) and unweighted Monte Carlo (right). After reweighting, the composition on Monte Carlo is by construction identical to the one observed in data. Data is shown sideband-subtracted.

trigger path	data	MC	weight
LOWPT	9%	33%	0.26
B_CHARM	27%	27%	1.01
HIGHPT	64%	39%	1.63

Table 6.1: Composition of trigger paths on B_s^0 data and Monte Carlo. The weights are obtained by dividing the fraction of a given trigger path in data by its fraction in Monte Carlo.

reweighting can be seen. Table 6.1 shows the data in numeric form, along with the weights that are calculated as the ratio between the fraction in data to the fraction in Monte Carlo. Events simulated in the B_CHARM scenario remain unchanged, the LOWPT scenario needs to be scaled down to approximately $\frac{1}{4}$ of its previous size, while the HIGHPT events are significantly more frequent on data than predicted in the simulation and are weighted up accordingly. This behaviour can be explained by the recent increase in Tevatron luminosity: The event acceptance requirements of the HIGHPT scenario are stricter than for the other two, so it can be operated at relatively high luminosities where the other scenarios are disabled by high prescales. Compared to the luminosity profile assumed in the trigger simulation, a larger share of data taking happens in high luminosity running conditions.

Neural Network Reweighting

Artificial neural networks have been used in classification problems throughout this thesis. Their task is to evaluate differences between the signal and background samples, reducing the multi-dimensionality of the problem to a single probability value for an event being signal or background. Finding differences between data and Monte Carlo can also be seen as a classification problem. No separation power is expected from the neural network if both samples are similar to each other, whereas disagreement between them will result in different outputs for the two samples.

Neural networks can not only be used to identify differences between the samples, but their output can also be used for reweighting one sample to make it similar to the other. NeuroBayes conveniently offers a neural network output which is a probability $P_{\text{signal}} \in [0 : 1]$ for having a signal event. Defining the real data to be signal and Monte Carlo to be background, the neural network output NN can be written as

$$NN = \frac{P_{\text{data}}}{P_{\text{data}} + P_{\text{MC}}},\tag{6.28}$$

from which the Monte Carlo weight

$$w_{\text{MC}} = \frac{P_{\text{data}}}{P_{\text{MC}}} = \frac{NN}{1 - NN}\tag{6.29}$$

follows.

The available B_s^0 dataset consists of roughly 2500 events. For reliably training a weighting neural network, about three to four times this sample size is necessary, as the danger of learning statistical fluctuations is high at small statistics. Therefore, the B_s^0 Monte Carlo in this tagger uses only the described opposite side and trigger reweighting. The B^0 and B^+ datasets are more than ten times as large as the B_s^0 sample, so the additional neural network reweighting is applied here. In this chapter, the B^0 case is shown and

discussed, the corresponding plots for the B^+ case can be found in appendix B.1.

Before the reweighting network is trained, the trigger composition is corrected as described in the last section. Nevertheless, the trigger path is also included as input variable for the network reweighting as a cross check. The other input quantities are the transverse momentum p_T, the pseudo-rapidity η and the angle ϕ of the reconstructed same side B meson. Equal sized samples of data (signal) and Monte Carlo (background) are used in the training process so that the data probability is 0.5 for events that are in good agreement.

In figure 6.9, the input variables and the neural network output for the B^0 event reweighting network are shown. The overall agreement between data and simulation in the selected variables is reasonable even before the network reweighting. The trigger composition does not show significant disagreement, which is expected because the Monte Carlo sample had already been reweighted before this comparison. A significant difference is visible in the B meson transverse momentum, with a tendency for higher p_T values on data compared to Monte Carlo. High p_T events will have high neural network output values, resulting in larger weights. In the pseudo-rapidity η, events in the very forward regions of the detector are more frequent on Monte Carlo than on data. These regions will be weighted down as a result of their relatively low neural network output. For the azimuthal angle ϕ, significant simulation problems are neither expected nor observed here, which is a reassuring result.

From the neural network output, the event weights are calculated according to equation 6.29. The weight to restore the correct opposite side B hadron composition, the trigger and the network wights are considered independent and therefore combined by multiplying all three. The resulting distribution of weights can be seen in figure 6.10, the average weight is one.

After applying these event weights to the Monte Carlo sample, the training process is repeated as a cross check. The observable differences should be significantly smaller after the reweighting. The variables and the neural network output of the validation round can be seen in figure 6.11. In comparison to the uncorrected situation shown in figure 6.9, the width of the neural network output distribution is visibly smaller, and the contributing variables do not show significant disagreement any more. The reweighting procedure can therefore be regarded as a successful improvement of the available Monte Carlo simulation.

6.3 Test With Other B Mesons

Studying the new flavour tagger on data is the key part in the validation procedure. Besides the relatively rare B_s^0 mesons, also its sister particles B^+ and B^0 , of which CDF has accumulated substantial amounts, can be used. Due to conservation of the electric charge, B^+ mesons are not subject to flavour oscillations. While this eliminates the need for flavour tagging in analyses, it opens up possibilities for tagging validation that
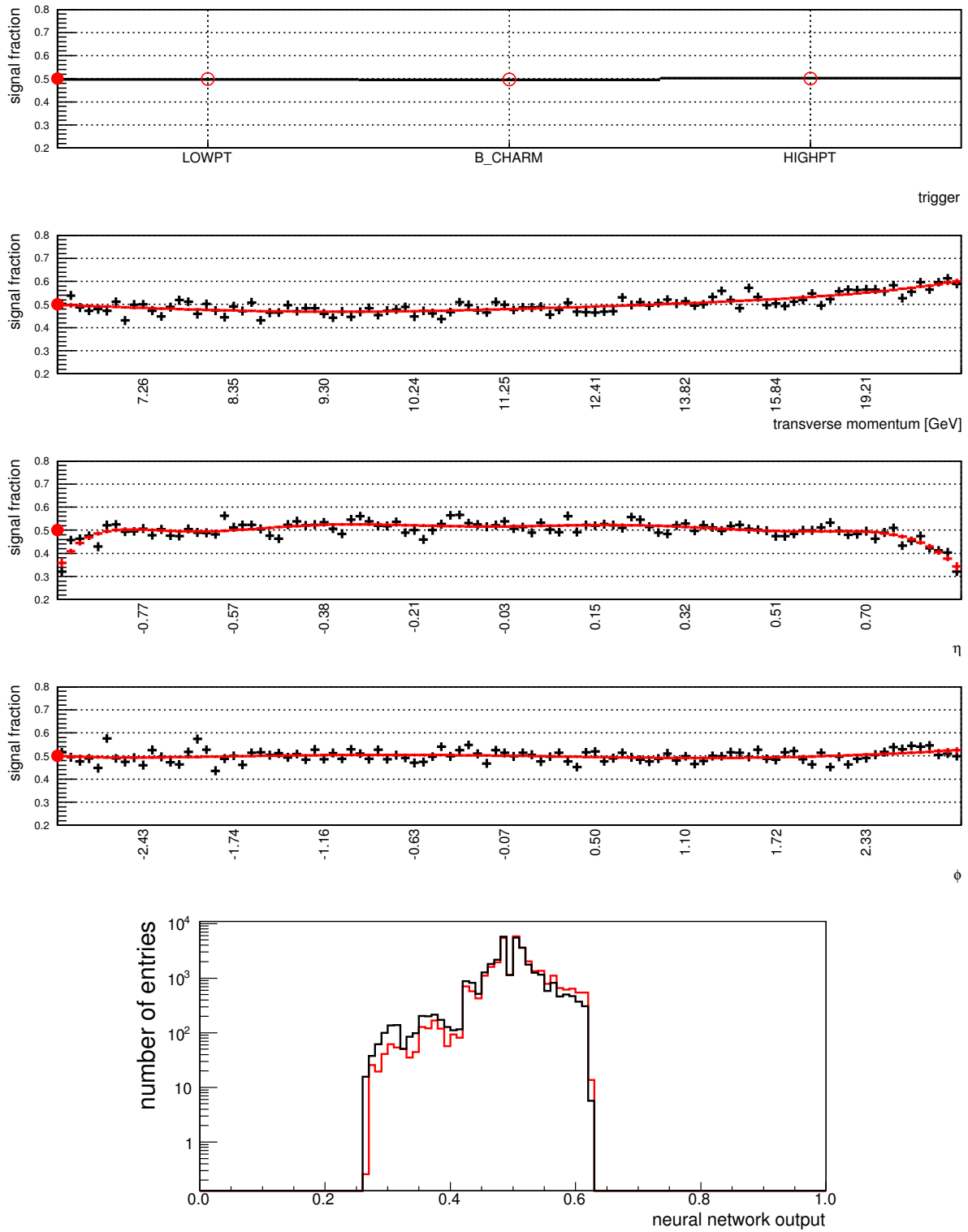
Figure 6.9: Input quantities and neural network output (on logarithmic scale) for the B^0 Monte Carlo reweighting network. The plotting conventions are explained in chapter 4.5.1.
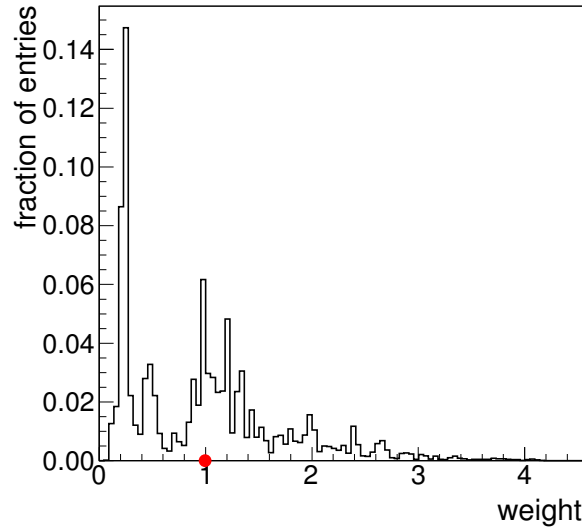
Figure 6.10: Distribution of event weights to correct differences between B^0 data and Monte Carlo. The mean value is indicated by the marker on the x axis.

are almost as straightforward as on Monte Carlo, because the true production flavour is always known. Concerning flavour tagging physics, B^0 mesons are very similar to B_s^0 mesons. Their considerably lower oscillation frequency removes some of the experimental challenges of the B_s^0 case, and due to their more frequent occurrence, larger samples allow for more precise measurements. In this section, validation procedures and results on both B^+ and B^0 mesons are discussed.

6.3.1 Charged $B^\pm$ Mesons

The production flavour of a reconstructed B^+ meson is unambiguously known because it is identical to its decay flavour. Validation of tagging performance can be done in a similar way as on Monte Carlo, with only minor complication from combinatorial background. The distribution of the different event level tagging quantities are known separately for B^+ and B^- mesons, and so also the relation between predicted and measured flavour can be obtained.

The physical aspects of B^+ mesons are considerably different from neutral B mesons, especially in same side tagging. As the initial $p\bar{p}$ state is charge neutral while the B^+ meson itself is charged, all other tracks have a higher probability for being oppositely charged. Because of this fundamental difference, B^+ data can not directly be used to validate the B_s^0 trained tagger. Instead a specialized B^+ version of the tagger was derived by retraining the *track flavour correlation networks* and the *combined tag network*. For

Figure 6.11: Validation of the B^0 Monte Carlo reweighting network with the reweighted Monte Carlo sample. The plotting conventions are explained in chapter 4.5.1.
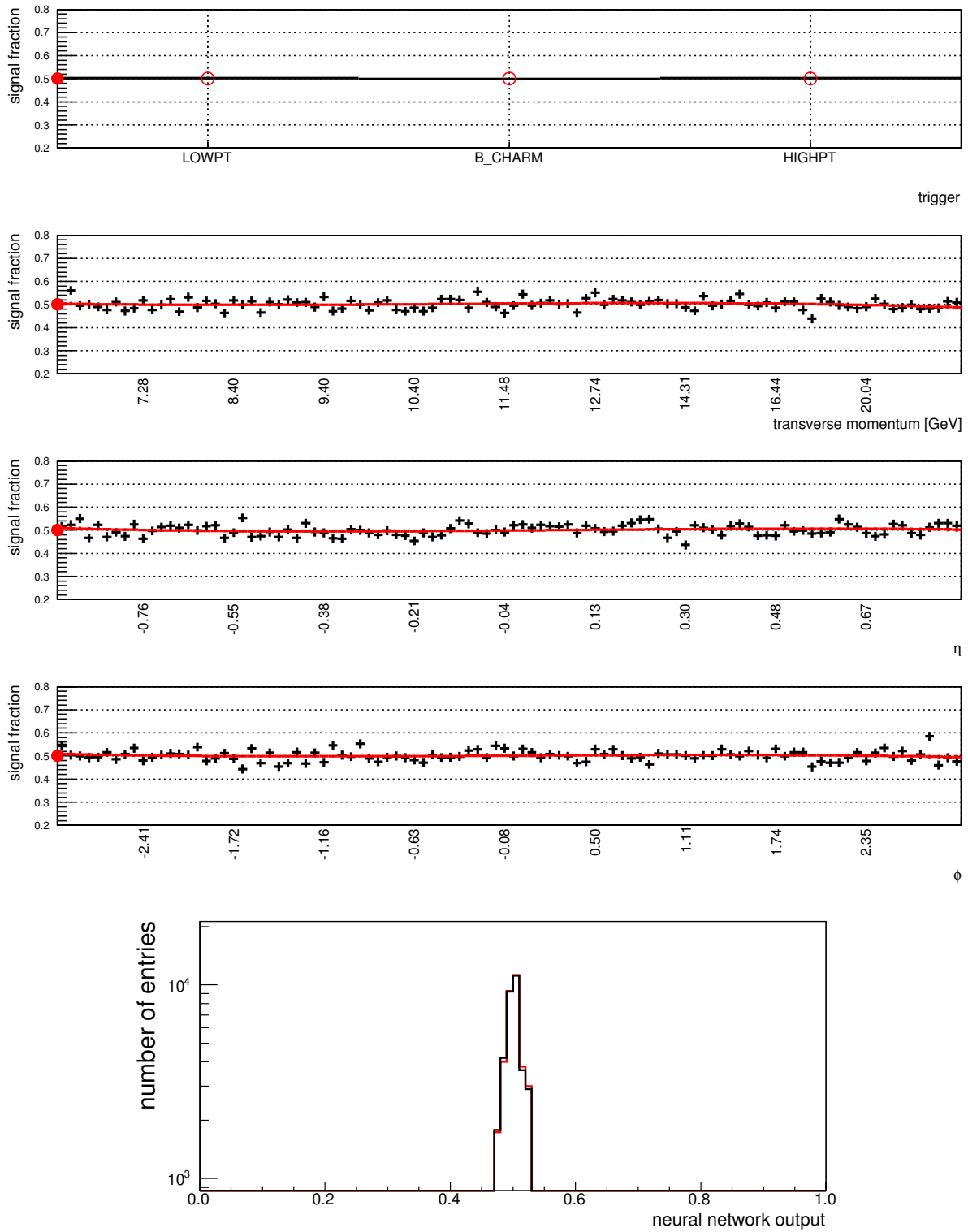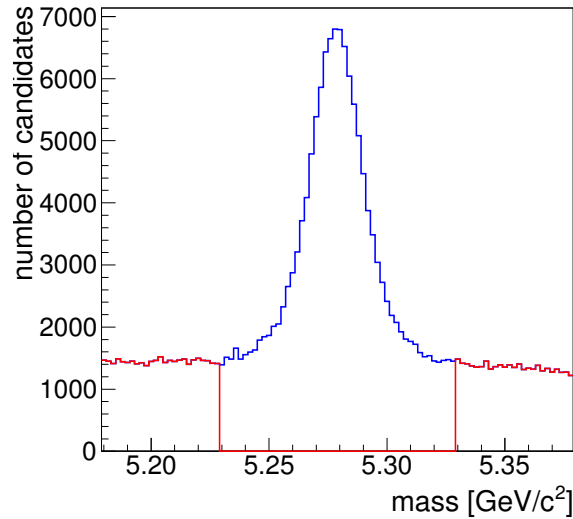
Figure 6.12: Mass spectrum of the B^+ validation sample. The blue peak region contains a mixture of signal and background candidates, the red sideband regions are pure background samples. Both sidebands together contain the same number of background candidates as expected in the peak region.

this, a B^+ Monte Carlo sample was generated in the same way as the B_s^0 Monte Carlo. While a flavour tagger for a charged meson has no application for analysis, it is here used for testing the tagging procedure itself along with the quality of the Monte Carlo simulation. Additional differences could arise from the fact that the B^+ data sample in the decay channel

$$B^{\pm} \rightarrow J/\psi \ K^{\pm}, \ J/\psi \rightarrow \mu^+\mu^- \tag{6.30}$$

was collected using a di-muon trigger, while the B_s^0 and B^0 are two-track trigger samples.

During the reconstruction of a B meson from its stable daughter particles, random combinations of tracks can form false candidates, called *combinatorial background*. Although advanced selection procedures with high purity and efficiency are available and in use, a signal sample will always contain a certain amount of background.

When studying the distribution of a given quantity on data, the influence of these background events can be taken into account by a *sideband subtraction*. The mass spectrum of reconstructed B meson candidates after candidate selection usually shows a distinct peak at the expected meson mass and more or less flat background regions. In the *signal region*, a mass window around the central peak value, a mixture of signal and background properties can be observed. The background contribution to the peak region can be accounted for by subtracting a pure background sample – taken from the *sideband region* outside the peak – from the distribution of the quantity of interest. In figure 6.12, the
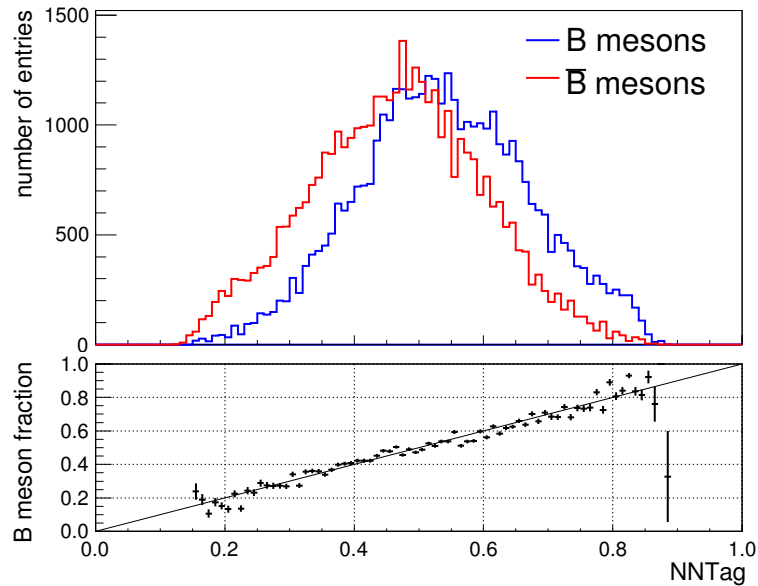
Figure 6.13: Output distribution of the combined neural network flavour tagger (NNTag) on sideband subtracted $B^{\pm} \to J/\psi K^{\pm}$ data. The true B meson fraction shown in the lower plot agrees well with the line indicating ideal behaviour.

mass spectrum of the B^+ data sample used for validation can be seen. Signal and sideband regions are indicated in the plot.

Figure 6.13 shows the final output of the B^+ version of the combined neural network tagger on real B^+ data. It can be seen from the lower plot that the relation between tagger output and true B meson probability is linear. It is also consistent with the indicated ideal situation, which means that the tagger output is a true B meson probability also on B^+ data. This is a very satisfying result and gives confidence in the quality of the Monte Carlo simulation as well as in the overall tagging strategy.

Also the ingredients in the final tag can be validated individually. The corresponding set of validation plots can be found in figure 6.14. All of them show consistency with the ideal behaviour, so no indications of problems is present in this check, providing confidence in the basic approach behind this tagger and the simulation.

6.3.2 B^0 Flavour Oscillations

The physics of flavour oscillation and tagging in the B^0 meson system is very similar to the B_s^0 case. The B^0 oscillation frequency is about 35 times slower than in the B_s^0 system, which poses significantly lower requirements on the resolution of the proper decay time measurement. Another notable difference to the B_s^0 case can be found in same side tagging: The leading fragmentation partner of the same side B meson is usually a kaon for B_s^0 and
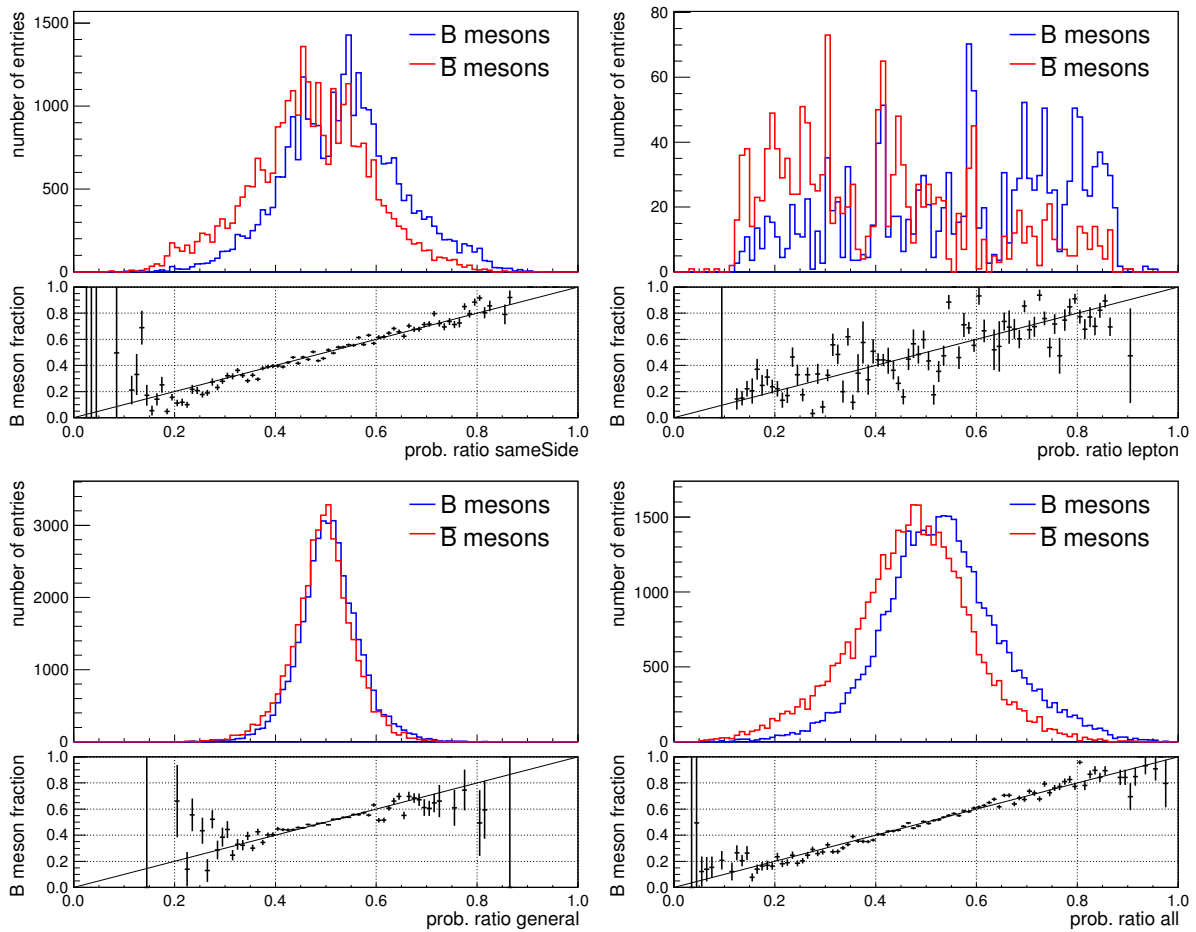
Figure 6.14: Distributions of event level tagging quantities constructed from the track flavour correlation probabilities on sideband subtracted $B^{\pm} \rightarrow J/\psi K^{\pm}$ data.
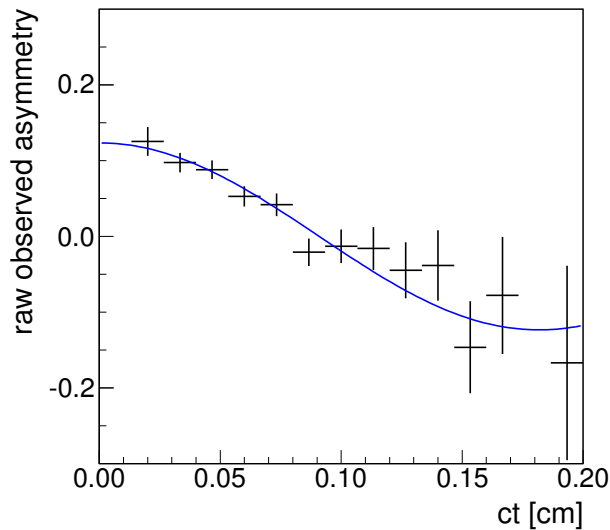
Figure 6.15: Raw observed B^0 oscillation asymmetry as a function of proper decay time, measured with the combined neural network tagger.

a pion for B^0. Nearly ten times as many reconstructed B^0 mesons are available in data compared to B_s^0 mesons. Furthermore, the B^0 oscillation frequency has been precisely measured by the B factories. All these reasons suggest that B^0 mesons provide a nearly ideal validation testbed.

Raw B^0 Oscillation Asymmetry

The easier experimental situation makes it possible to directly plot the time dependent B^0 oscillation asymmetry and thereby deliver a visual proof that the new flavour tagger is able to measure flavour oscillations.

Figure 6.15 shows the asymmetry (see equation 2.19) between *mixed* (production and decay flavour are opposite) and *unmixed* (production and decay flavour are equal) B^0 mesons as determined by the B^0 tagger as a function of proper decay time. Only binary tagging decisions were used, ignoring the predicted dilution delivered by the tagger. We fit the asymmetry with a function

$$A(t) = A_0 \cdot \cos(\Delta m_d\, t), \tag{6.31}$$

which is the B^0 equivalent of equation 2.22. The parameter values

$$A_0 \;=\; 0.12 \pm 0.01, \tag{6.32}$$
$$\Delta m_d \;=\; (0.518 \pm 0.023)\; \hbar\, \mathrm{ps}^{-1}, \tag{6.33}$$
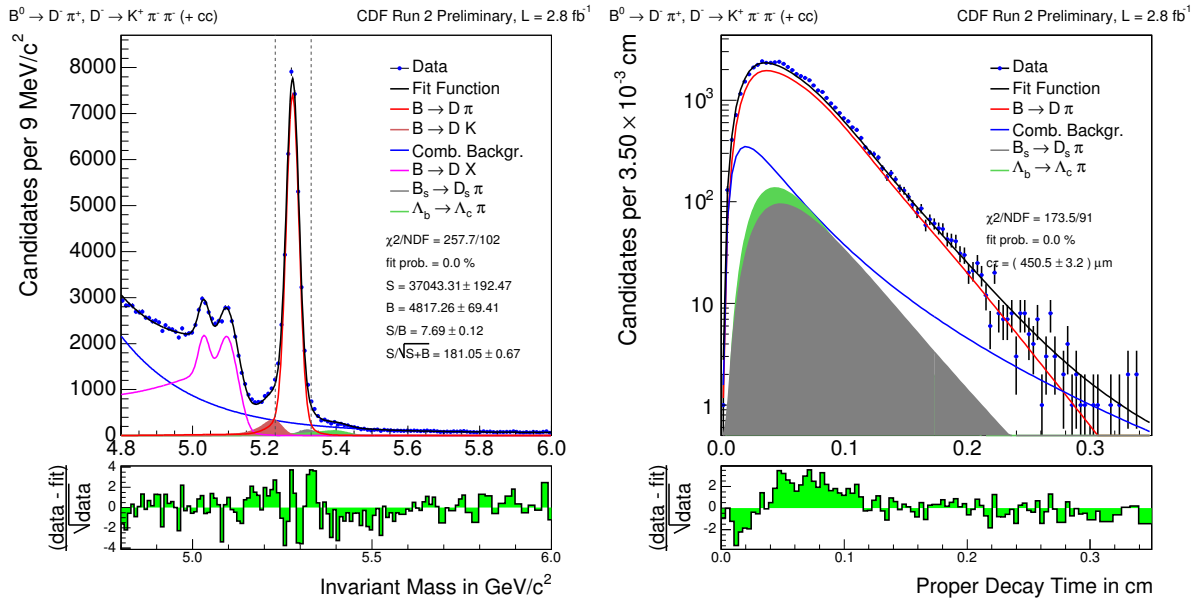
Figure 6.16: Fit projections in mass (left) and proper decay time (right), using the B^0 decay $B^0 \to D^-\pi^+, D^- \to K^+\pi^-\pi^-$.

have been determined in the fit. The measured oscillation frequency is in good agreement with the world average value $\Delta m_d = (0.508 \pm 0.005)\ \hbar\ \mathrm{ps}^{-1}$ [14]. While the physical amplitude of this asymmetry is 1, the lower value observed for A_0 is caused mainly by wrong tagging decisions.

Full Mixing Fit

The full B^0 mixing analysis is very similar to the measurement of B_s^0 oscillations. In an unbinned maximum likelihood fit, the oscillation frequency and the *observed oscillation amplitude* $\mathcal{A}$ are determined. For a correctly predicted dilution, $\mathcal{A}$ is one, while underestimation of the dilution results in larger and overestimation in lower values of the amplitude. An overview on the mixing measurement is given in section 2.2. As it is an extensive analysis on its own, it is work in progress in the context of another thesis [26] within the Karlsruhe CDF group. Preliminary results interesting for tagger validation are presented here.

In figure 6.16, the fit projections to invariant mass and proper decay time are shown. Both spectra are well described by the fit parametrization, indicating a thorough understanding of all background contributions found in the dataset. The resulting oscillation frequency

$$\Delta m_d = (0.484 \pm 0.034)\ \hbar\ \mathrm{ps}^{-1} \tag{6.34}$$
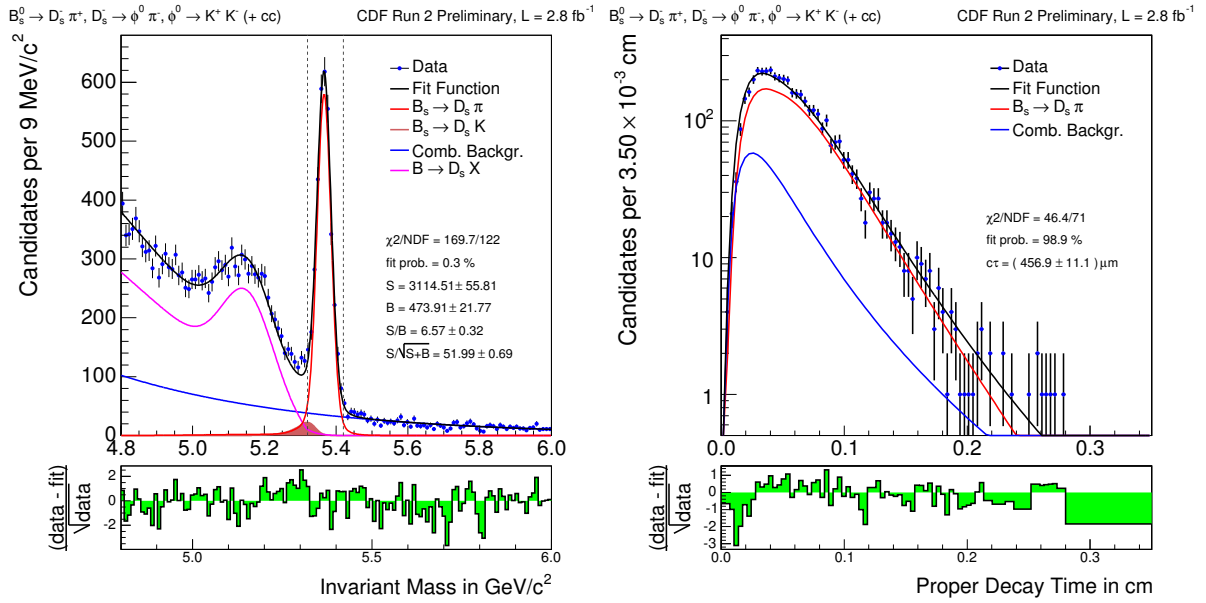
Figure 6.17: Fit projections in mass (left) and proper decay time (right) in the decay channel $B_s^0 \rightarrow D_s^\pm \pi^\mp$, $D_s^\pm \rightarrow \phi \pi^\pm$, $\phi \rightarrow K^+ K^-$ that was used for the validation analysis.

agrees well with the world average value $\Delta m_d = (0.508 \pm 0.005)$ $\hbar$ ps^{-1} [14]. Also the observed amplitude

$$\mathcal{A} = 0.93 \pm 0.08 \qquad (6.35)$$

is consistent with one. Both values are in good agreement with the expectation and give credibility to the predicted tagging performance.

6.4 B_s Flavour Oscillations

The B_s^0 oscillation measurement is one of the main applications of this flavour tagger, while serving as validation opportunity at the same time. The B_s^0 mixing signal has been observed by the CDF experiment three years ago [5] and the oscillation frequency has been measured to be $\Delta m_s = (17.77 \pm 0.10)$ $\hbar$ ps^{-1}. Once an oscillation signal can be observed, the measured amplitude serves as a consistency check for the predicted dilution. The observed oscillation amplitude $\mathcal{A}$ is one if the dilution is well predicted.

In the published CDF analysis, multiple hadronic and semileptonic B_s^0 decay channels have been combined. In the meantime, a significantly larger data sample has been accumulated, allowing a preliminary check on the decay channel

$$B_s^0 \rightarrow D_s^\pm \pi^\mp, \ D_s^\pm \rightarrow \phi \pi^\pm, \ \phi \rightarrow K^+ K^-. \qquad (6.36)$$

alone. The new neural network B flavour tagger presented in this thesis has been used to perform the full B_s^0 mixing analysis in this decay channel for validation. Fit projections for mass and proper decay time can be seen in figure 6.17, underlining the thorough understanding of dataset composition and the maturity of the mixing analysis. The measured oscillation frequency

$$\Delta m_s = (17.65 \pm 0.14) \ \hbar \ \mathrm{ps}^{-1} \tag{6.37}$$

is consistent with the previously measured value, and also the observed oscillation amplitude

$$\mathcal{A} = 0.98 \pm 0.28 \tag{6.38}$$

is consistent with one. The tagging power on B_s^0 data

$$\mathcal{T} = 4.6\% \tag{6.39}$$

has been determined from the predicted dilution and corrected by the observed oscillation amplitude. It is consistent with the tagging power measured on the validation Monte Carlo sample.

The observed oscillation amplitude confirms the validity of the dilution prediction, which also implies that the Monte Carlo simulation is sufficiently well under control and has an acceptable agreement with real data. Together with the encouraging cross check results from the B^+ meson studies and the B^0 oscillation measurement, this result is another indication that the new flavour tagger is ready for application. The update of the B_s^0 oscillation measurement using all available B_s^0 decay channels will be the final validation and the first application for the tagger, so that studies of CP violation in the B_s^0 meson system can benefit from improved flavour tagging.

6.5 Outlook: Particle Identification

The reconstruction of a charged particle track based on hit locations in the CDF tracking system allows for a precise momentum measurement, but spatial information alone is not enough to determine the type of particle associated with a given track. An estimate on the mass of a particle can be calculated by combining the momentum measurement with velocity information from the time of flight detector (TOF) or with the specific energy loss dE/dx in the Central Outer Tracker. In the context of flavour tagging, this is especially useful to distinguish kaons from pions, as kaons play an important role in B_s^0 same side tagging.

Despite the usefulness of particle identification information (PID), it can not be used without complications. Providing a sufficiently accurate simulation is a challenging task and can up to now only be achieved by parametrizing the observed behaviour in data.

As the performance and availability of the relevant detector components is a complicated function of various aspects, a data based calibration of both TOF and dE/dx quantities must be established before they can be used in analysis. These calibrations are often not available for a considerable time after data taking, delaying the application in every analysis that depends on them.

These limitations have proven impedimental during the development work on this flavour tagger, especially concerning the agreement between data and simulation. Therefore no TOF or dE/dx information is used in the current version of the new flavour tagger, making it ready for application on future new data without waiting for calibrations. After the agreement between data and simulation for the PID-free version of the tagger are understood and its development is finished, particle identification can be included again to benefit from kaon/pion separation in B_s^0 same side tagging.

A study of the possible gain by particle identification for the new tagger is presented in this section. It was done by including TOF and dE/dx information based on the existing simulation for the first 1 fb^{-1} in the track flavour correlation networks of the B_s^0 tagger. The additional PID input quantities for the same side correlation network are pion and kaon probabilities from TOF, dE/dx and from a likelihood based combination of both (CLL). The track based same side correlation network and the event level combined tag network have been retrained and applied on a Monte Carlo sample. The simulation is not valid for later data, and good agreement between data and simulation is not expected for this study.

The three most important input variables for the modified same side correlation network can be seen in figure 6.18. In comparison to the original network (see figure 5.11), the role of p_L along the B meson direction has been taken over by the combined particle ID likelihood ratio (CLL) for kaon hypothesis, which is the most significant input variable now. High kaon probabilities correspond to strong correlations between track charge and B flavour, which is expected for kaons being the leading fragmentation particles for B_s^0 mesons. The second and third most input variables have swapped their position, but remain unchanged otherwise.

Figure 6.19 shows the broader neural network output distribution corresponding to increased separation power of the PID-version in comparison to the original network output distribution. The output distributions of both versions of the combined tag network are shown in figure 6.20. The increased separation power leads to more events in the tail regions of the output distribution.

In the same way as described in section 6.1, the tagging power of the PID-version is determined on the validation Monte Carlo sample. The determined value of

$$\mathcal{T} = \epsilon D^2 = 6.7\% \tag{6.40}$$

corresponds to a 40% increase compared to $\mathcal{T} = \epsilon D^2 = 4.6\%$ of the version without particle identification. This confirms the potential for improvement with the help of TOF
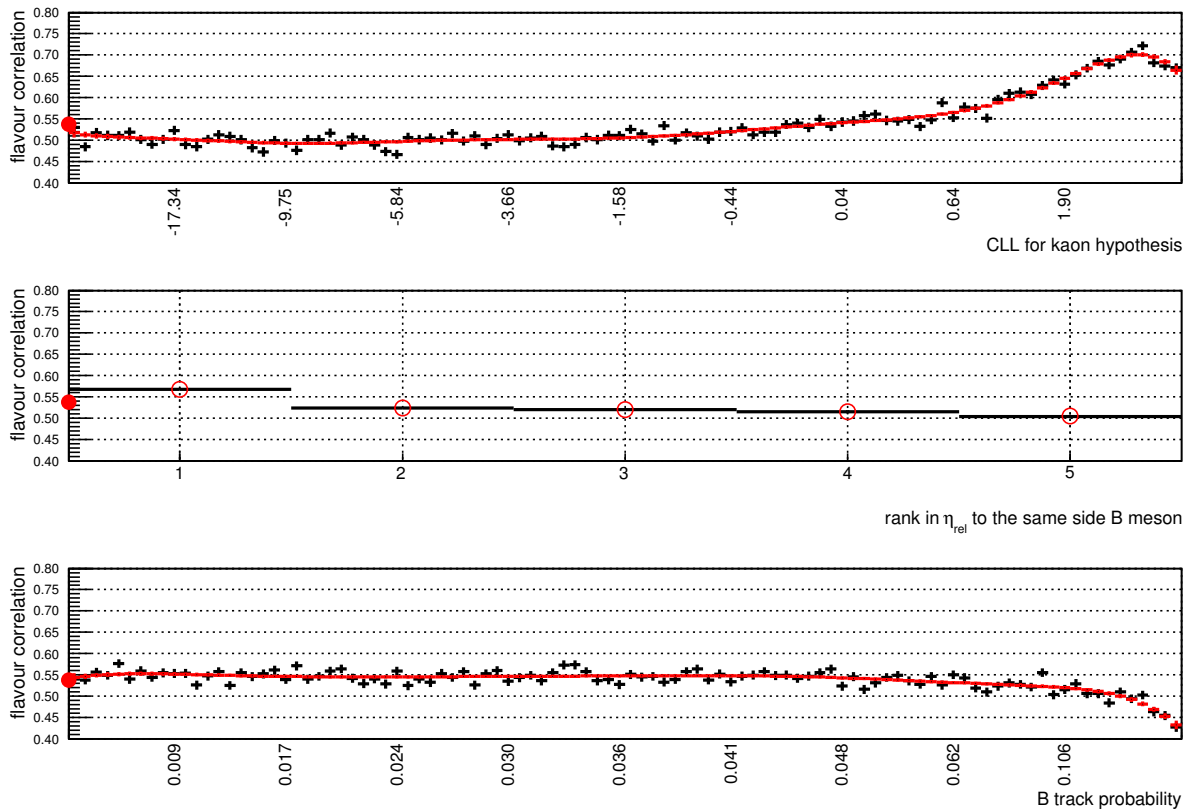
Figure 6.18: The three most important input quantities for the same side correlation network, including particle identification information. The plotting conventions are explained in chapter 4.5.1.
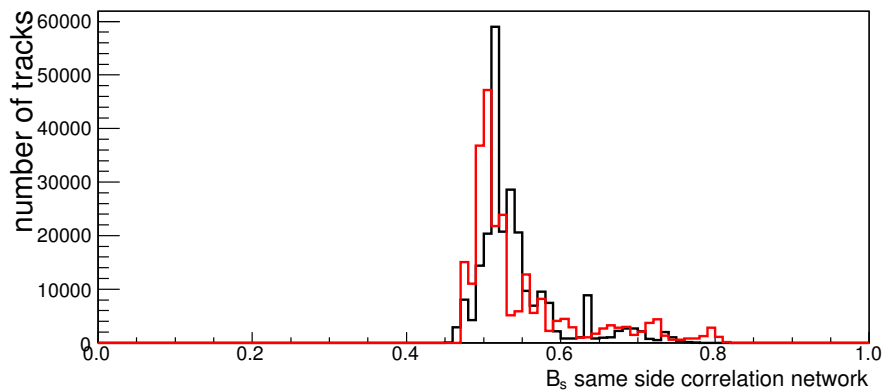
Figure 6.19: Neural network output distribution of the B_s^0 same side correlation network without (black) and with (red) particle identification variables.
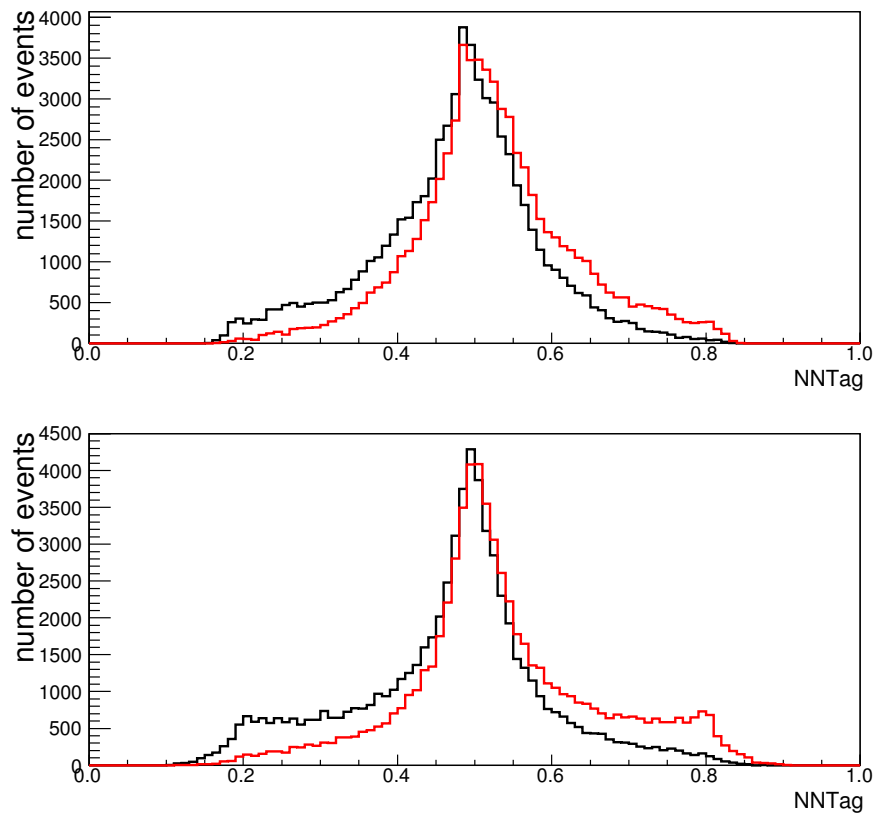
Figure 6.20: Neural network output distribution of the B_s^0 combined tag network without (top) and with (bottom) particle identification in same side tagging on Monte Carlo. True B mesons are shown in red, true $\bar{B}$ mesons in black. The network with particle identification is able to enrich the fraction of well-tagged events in the tails of the distributions.

and dE/dx information. As soon as calibrations for the latest data and a sufficiently accurate simulation are available, the power of the new combined flavour tagger can be significantly improved as was shown in this study.

7 Conclusion

Flavour tagging is a central element of all B physics analyses that involve resolving time dependent neutral B meson oscillations, most prominent among them the measurement of the B_s^0 oscillation frequency. Its goal is to determine whether a meson was produced as B or $\bar{B}$ meson. In this thesis, an advanced approach to B flavour tagging has been presented. Flavour information from all available sources is extracted in a uniform procedure and combined into a probability statement on the production flavour. Existing CDF flavour taggers are focused on individual information sources and can therefore not benefit from correlations between them.

The new flavour tagger consists of three main parts. In a first step, quantities useful for the later stages are obtained and an inclusive reconstruction of the second B hadron in an event is done. The second step is the heart of the procedure, where a probability for a correlation between track charge and the B meson production state is gained for every track. In the third step, these probability values of all tracks are combined so that an event level flavour tag is obtained. Artificial neural networks are used in all three parts to benefit from good classification performance. The tagger has been developed on simulated B_s^0 events and is now implemented within the CDF software framework.

The figure of merit for flavour tagging is the tagging power $\mathcal{T} = \epsilon D^2$ with *efficiency* ϵ and *dilution* D. The efficiency corresponds to the fraction of events which have been given a tagging decision at all, while the dilution is a measure of confidence associated with a decision given by the tagger. Despite its name, a high dilution indicates a more reliable decision and is favourable for a tagger. The relation between the dilution D and the probability P to have a correct tag is given by $D = 2 \cdot P - 1$. Having suboptimal tagging decisions results in an effective reduction of the available sample of B mesons. The combination ϵD^2 indicates the size of a hypothetical sample with perfect tagging decisions that has the same overall statistical significance as the reference sample.

The new flavour tagging algorithm presented here provides a tagging decision for all events that have at least one additional track not used in the exclusive reconstruction of the studied B meson. This is the case for more than 99% of all events, so the efficiency is virtually one. The overall tagging power on simulated B_s^0 events has been determined to be

$$\mathcal{T} = 4.6\%. \tag{7.1}$$

Because the agreement between simulated events and real data is always limited from fundamental reasons, the behaviour of the flavour tagger needs to be verified on data.

This was done on experimentally favourable B^+ and B^0 mesons as well as in the B_s^0 oscillation measurement. Dedicated B^+ and B^0 versions of the tagger have been derived by training the relevant neural networks on corresponding samples of simulated events. Validation results on B^+ mesons show good agreement between data and simulation, giving confidence in the simulation quality and the validity of the tagging procedure. The analysis of B^0 oscillations is very close to the B_s^0 case on the physics side, but has the experimental advantages of higher statistics and a lower oscillation frequency. The observed oscillation amplitude on data is consistent with the expected value of 1, and the frequency is consistent with the world average value, indicating good understanding of both simulation and the tagger itself. Also preliminary results of the B_s^0 oscillation analysis with the new tagger on real data show good agreement between simulation and application, although with a larger statistical uncertainty than in the B^0 case due to the experimental challenges in the B_s^0 system.

To eliminate a possible source of simulation problems and because calibrations are often not available for a considerable time after data taking, no information from the specific energy loss and time of flight system for particle identification is used in the current version of the flavour tagger. As improved kaon identification would especially help in B_s^0 same side tagging, the gain in stability and robustness is bought dearly by sacrificing overall tagging power. To quantify the cost of this tradeoff, the existing simulation of particle identification has been included in a testing version, disregarding the disagreement to data. In this study, the tagging power has been determined to be $\mathcal{T} = 6.7\%$, which is a 40% increase compared to the current version without particle identification. It is therefore highly desirable to overcome the technical difficulties with particle identification and include it in future versions of the flavour tagger.

Up to now, the CDF collaboration made use of a neural network combined opposite side tagger with a tagging power of $\mathcal{T} = 1.8\%$ and of a same side kaon tagger yielding $\mathcal{T} = 3.7\%$, where the latter depends on information for particle identification. The combined tagging power of both is not stated, but can not be higher than the sum of the two. Thus the new tagger presented in this thesis offers a significant increase in tagging power for data that has no calibrated particle identification available. When this calibration along with a reliable simulation is available, the new tagger can reach a prospective tagging power that is about 20% higher than that of the existing taggers combined.

The B_s^0 oscillation analysis with the increased CDF dataset since the last published result will be the first major application and constitute the final validation for the new flavour tagger. Afterwards, the search for CP violation in $B_s^0 \to J/\psi\,\phi$, in $B_s^0 \to J/\psi\,f_0(980)$ and in $B_s^0 \to K^+ K^-$ decays, which are sensitive to New Physics effects, will benefit from the new flavour tagger.

A Neural Network Input Variables

Key to read the tables:

- *rank*: importance rank for the given variable

- *index*: variable number, starting with two because of NeroBayes numbering conventions

- *added*: additional significance in σ added by this variable, given the higher ranked variables are already included

- *only this*: significance in σ for only this variable

- *loss*: lost significance in σ when this variable is removed from the network with all others remaining

- *corr.*: global correlation of this variable to the other variables

- *identifier*: variable name in short form

- a separating line in the table indicates which variables have been automatically removed because their contribution was not above the required precut

A.1 First B Track Network

rank	index	added	only this	loss	corr.	identifier
1	21	258.08	258.08	67.98	82.2%	jetData_d0signi
2	2	110.50	118.00	36.37	89.0%	stdData_pt
3	10	91.97	217.97	66.93	78.1%	bMesonData_d0
4	11	82.39	116.12	43.47	95.2%	bMesonData_deltaPhi
5	13	66.05	66.78	42.13	80.7%	nOtherTracks
6	8	59.05	89.25	58.51	14.5%	muonProbFlag
7	6	56.78	15.00	52.52	27.7%	nSiHitsL00
8	4	37.67	258.92	36.75	87.8%	stdData_d0signi
9	7	37.82	18.15	32.10	56.9%	stdData_eta
10	12	23.78	99.71	21.98	95.3%	bMesonData_deltaR
11	14	22.12	53.28	25.71	78.7%	nTracksJet
12	5	20.56	33.32	20.74	33.4%	stdData_z0signi
13	9	14.29	12.22	14.74	29.8%	bMeson_pt
14	22	9.65	53.09	15.63	77.3%	inSameSideJet
15	20	14.56	62.81	16.68	76.6%	jetData_pt_ptJet
16	16	8.86	14.85	12.58	54.4%	jetData_ptRelFlag
17	18	10.04	70.17	14.80	86.8%	jetData_rapidity
18	15	9.77	30.02	9.53	32.2%	jetData_ptJet
19	3	8.39	121.35	10.51	88.7%	stdData_ptRank
20	19	8.56	96.21	8.23	90.5%	jetData_rapidityRank
21	17	7.05	38.98	7.05	48.1%	jetData_ptRel

A.2 Secondary Vertex Network

rank	index	added	only this	loss	corr.	identifier
1	3	288.65	288.65	126.67	72.0%	bProb2
2	12	104.20	214.31	18.30	87.7%	lxySigni
3	6	91.14	143.45	51.85	54.9%	inSameJet
4	15	56.60	37.26	52.41	23.6%	fitChi2
5	2	60.38	235.36	54.02	72.7%	bProb1
6	13	42.49	189.72	22.52	76.0%	abs_deltaPhi
7	14	29.31	161.21	28.31	50.6%	abs_deltaEta
8	9	26.87	41.19	29.80	34.0%	cosThetaStar
9	10	17.64	218.31	19.38	89.8%	lxy
10	8	16.77	113.70	16.54	45.5%	mass
11	5	16.35	78.70	10.63	60.4%	bProbRank2
12	7	15.12	29.26	15.13	5.5%	abs_charge
13	4	6.45	58.84	6.50	62.9%	bProbRank1
14	11	4.23	41.03	4.23	59.1%	sigmaLxy

A.3 Second B Track Network

rank	index	added	only this	loss	corr.	identifier
1	2	537.61	537.61	341.40	67.3%	bTrackProb1
2	12	62.92	163.57	44.24	86.4%	vertexData_d0signi
3	3	94.92	201.20	61.30	65.2%	secVertex_bProb
4	9	78.74	303.28	28.93	87.4%	vertexData_cosThetaRest
5	20	39.22	85.71	40.48	11.2%	stdData_isEfromBcand
6	4	35.61	102.59	39.57	63.0%	secVertex_lxy
7	5	35.30	303.00	43.46	76.0%	vertexData_usedInFit
8	8	33.01	284.56	15.11	89.4%	vertexData_pRest
9	10	23.05	234.63	14.74	88.6%	vertexData_d0
10	6	17.86	268.02	16.26	76.8%	vertexData_inJet
11	11	11.36	258.62	12.19	73.5%	vertexData_sigmaD0
12	16	10.18	167.31	16.64	71.7%	stdData_pt
13	7	12.67	116.80	12.89	54.4%	vertexData_mass
14	21	11.60	20.29	11.56	3.4%	stdData_electron_noconv
15	13	10.47	54.96	10.62	41.0%	vertexData_z0
16	19	10.01	20.90	10.08	3.8%	stdData_muonProb
17	18	9.46	130.21	9.49	26.4%	stdData_muonProbFlag
18	14	6.31	109.25	9.63	67.9%	vertexData_sigmaZ0
19	17	9.24	33.60	9.25	56.4%	stdData_eta
20	15	0.71	22.71	0.71	17.7%	vertexData_z0signi

A.4 Lepton from B Network

rank	index	added	only this	loss	corr.	identifier
1	2	94.74	94.74	43.68	79.6%	bTrackProb2
2	9	29.10	60.92	26.90	48.8%	bMesonData_pTrel
3	10	15.47	40.50	14.11	38.1%	bMesonData_pLrel
4	13	12.48	48.42	16.35	56.1%	vertexData_pRest
5	12	11.04	34.51	8.53	82.2%	vertexData_inJet
6	7	8.70	3.83	8.61	23.9%	stdData_eta
7	4	7.42	47.73	7.31	51.1%	electronAndNotConversion
8	8	5.61	31.90	5.38	39.9%	bMesonData_d0
9	14	3.62	50.07	4.54	82.9%	vertexData_cosThetaRest
10	11	2.95	48.97	2.75	85.0%	vertexData_usedInFit
11	3	1.56	35.03	1.35	46.6%	isEfromBcand
12	6	1.19	28.66	1.19	39.0%	muonProb
13	5	0.00	35.04	0.00	100.0%	isMufromBcand

A.5 Same Side Correlation Network

rank	index	added	only this	loss	corr.	identifier
1	23	23.60	23.60	11.53	83.9%	bMesonData_pLrel
2	2	23.12	19.46	9.97	67.9%	bTrackProb2
3	22	18.61	23.64	2.76	94.6%	bMesonData_etaRelRank
4	10	9.73	21.61	3.23	99.1%	jetData_d0
5	25	9.99	9.54	6.45	67.8%	bMesonData_pTrel
6	24	6.46	25.67	5.91	69.4%	bMesonData_pLrelRank
7	19	6.00	20.78	5.63	97.9%	bMesonData_deltaR
8	21	5.91	20.07	6.56	97.5%	bMesonData_etaRel
9	26	5.18	4.90	4.75	76.9%	bMesonData_pTrelRank
10	14	4.22	22.16	2.84	82.5%	jetData_rapidity
11	3	4.01	5.89	3.97	11.3%	lepFromBProb
12	13	2.83	8.51	2.86	67.7%	jetData_ptRel
13	5	2.81	2.87	2.72	30.0%	vertexData_z0
14	15	2.55	17.80	2.99	62.9%	jetData_rapidityRank
15	8	2.13	9.97	2.10	83.7%	vertexData_cosThetaRest
16	6	2.10	10.16	2.30	77.2%	vertexData_inJet
17	16	1.73	21.44	1.76	99.1%	bMesonData_d0
18	17	1.56	11.84	0.78	81.0%	bMesonData_deltaEta
19	4	1.30	4.74	1.33	20.4%	vertexData_d0
20	20	0.86	23.33	0.87	94.7%	bMesonData_deltaRRank
21	7	0.84	17.18	0.83	81.9%	vertexData_pRest
22	18	0.48	14.47	0.48	84.9%	bMesonData_deltaPhi
23	9	0.25	2.80	0.25	30.2%	jetData_nTracksJet
24	12	0.00	0.00	0.00	0.0%	jetData_inSameSideJet
25	11	0.00	0.00	0.00	0.0%	jetData_inOppositeSideJet

A.6 Lepton Correlation Network

rank	index	added	only this	loss	corr.	identifier
1	9	20.20	20.20	12.57	65.7%	jetData_nTracksJet
2	13	22.04	16.63	11.53	66.3%	jetData_ptRel
3	8	17.46	19.90	11.13	48.8%	vertexData_cosThetaRest
4	16	13.65	16.48	4.92	90.0%	bMesonData_d0
5	7	12.01	19.31	9.16	56.4%	vertexData_pRest
6	25	7.80	19.45	3.27	79.1%	bMesonData_pTrel
7	23	5.56	16.10	3.31	70.0%	bMesonData_pLrel
8	5	5.08	8.79	4.16	43.3%	vertexData_z0
9	15	4.32	15.50	2.78	69.9%	jetData_rapidityRank
10	2	3.85	16.67	1.22	89.5%	bTrackProb2
11	20	1.98	9.55	2.36	84.6%	bMesonData_deltaRRank
12	24	2.49	7.09	2.51	77.0%	bMesonData_pLrelRank
13	18	1.44	4.38	2.35	70.5%	bMesonData_deltaPhi
14	19	1.80	3.37	2.31	81.9%	bMesonData_deltaR
15	21	2.08	3.21	2.14	71.3%	bMesonData_etaRel
16	11	1.64	4.08	1.77	46.0%	jetData_inOppositeSideJet
17	4	1.26	5.36	1.66	65.6%	vertexData_d0
18	3	1.26	23.17	1.27	93.8%	lepFromBProb
19	6	1.07	1.33	1.05	65.5%	vertexData_inJet
20	17	0.90	1.87	0.95	32.9%	bMesonData_deltaEta
21	26	0.84	1.19	0.81	33.1%	bMesonData_pTrelRank
22	14	0.81	13.45	0.80	74.9%	jetData_rapidity
23	10	0.42	15.07	0.42	90.5%	jetData_d0
24	22	0.02	8.97	0.02	85.9%	bMesonData_etaRelRank
25	12	0.00	0.00	0.00	0.0%	jetData_inSameSideJet

A.7 General Correlation Network

rank	index	added	only this	loss	corr.	identifier
1	2	10.79	10.79	4.85	67.5%	bTrackProb2
2	13	5.80	6.35	4.95	55.3%	jetData_ptRel
3	14	5.58	5.44	2.10	68.8%	jetData_rapidity
4	19	5.18	6.51	2.52	80.6%	bMesonData_deltaR
5	8	4.05	8.72	4.11	56.2%	vertexData_cosThetaRest
6	16	2.87	6.04	2.08	68.5%	bMesonData_d0
7	17	2.60	2.33	2.58	10.4%	bMesonData_deltaEta
8	9	2.35	4.48	1.44	55.2%	jetData_nTracksJet
9	7	2.60	3.27	1.96	63.9%	vertexData_pRest
10	18	2.30	5.93	2.34	53.0%	bMesonData_deltaPhi
11	20	2.05	3.90	1.94	71.0%	bMesonData_deltaRRank
12	15	1.36	5.59	1.16	67.3%	jetData_rapidityRank
13	26	1.29	3.56	0.93	48.2%	bMesonData_pTrelRank
14	4	1.33	4.88	1.54	40.5%	vertexData_d0
15	6	1.20	4.25	1.01	67.0%	vertexData_inJet
16	10	0.84	5.34	0.86	69.8%	jetData_d0
17	25	0.86	7.95	0.92	76.0%	bMesonData_pTrel
18	22	0.77	2.89	1.00	78.7%	bMesonData_etaRelRank
19	24	0.73	2.09	0.80	70.4%	bMesonData_pLrelRank
20	5	0.71	0.99	0.69	25.1%	vertexData_z0
21	11	0.58	3.11	0.61	53.0%	jetData_inOppositeSideJet
22	21	0.46	5.38	0.46	83.3%	bMesonData_etaRel
23	23	0.26	6.88	0.26	54.6%	bMesonData_pLrel
24	12	0.00	0.00	0.00	0.0%	jetData_inSameSideJet
25	3	0.00	0.00	0.00	0.0%	lepFromBProb

A.8 Combined Tag Network

rank	index	added	only this	loss	corr.	identifier
1	2	76.84	76.84	5.55	98.8%	lhrAll
2	6	13.98	41.03	16.60	62.2%	vertexCharge
3	3	7.68	59.75	6.08	98.1%	lhrSameSide
4	4	6.96	36.09	4.74	94.4%	lhrLepton
5	5	1.70	32.59	1.70	94.5%	lhrGeneral

B Data / Monte Carlo agreement
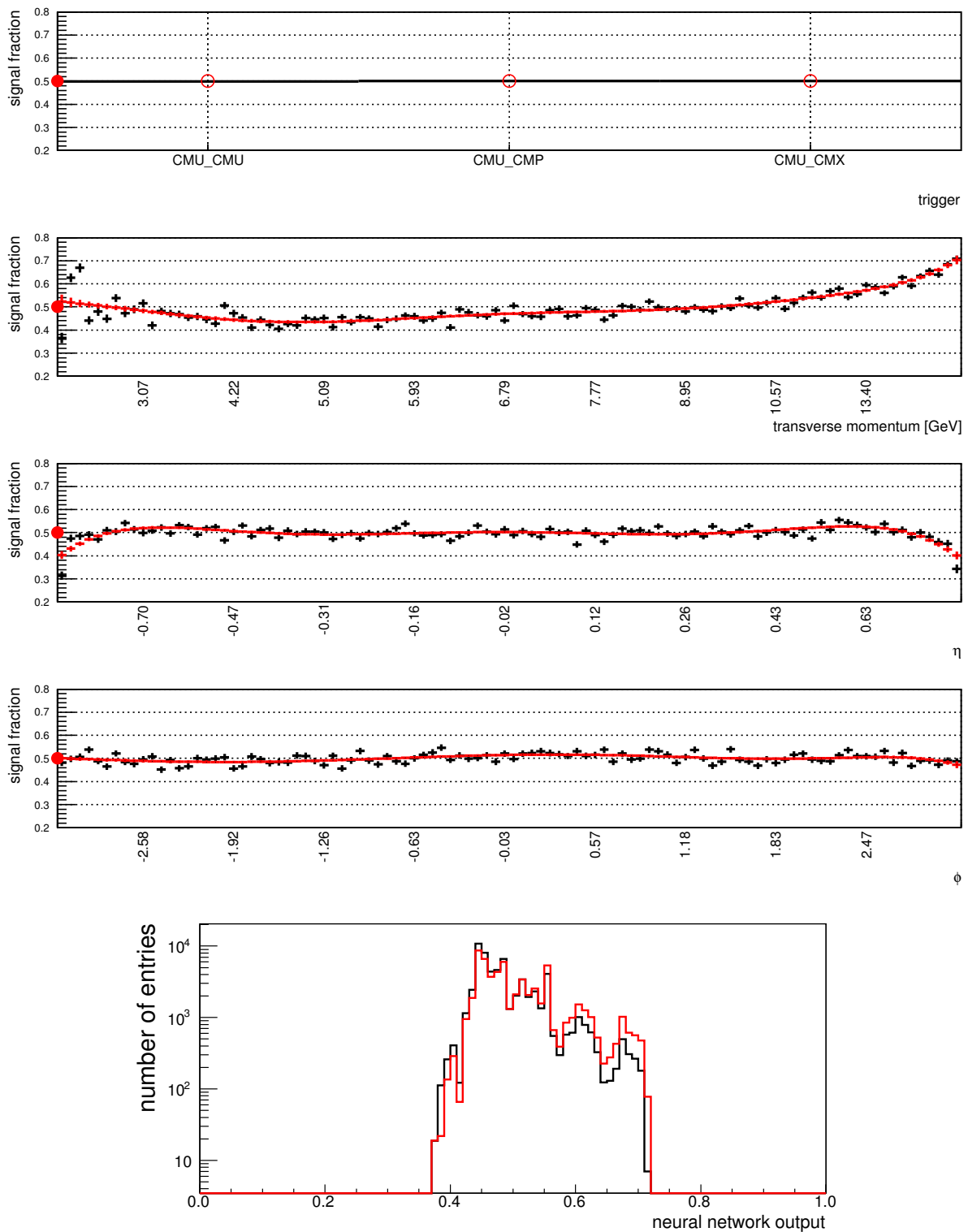
B.1 Reweighting B^+ Monte Carlo

Figure B.1: Input quantities and neural network output (on logarithmic scale) for the B^+ Monte Carlo reweighting network. The plotting conventions are explained in chapter 4.5.1.
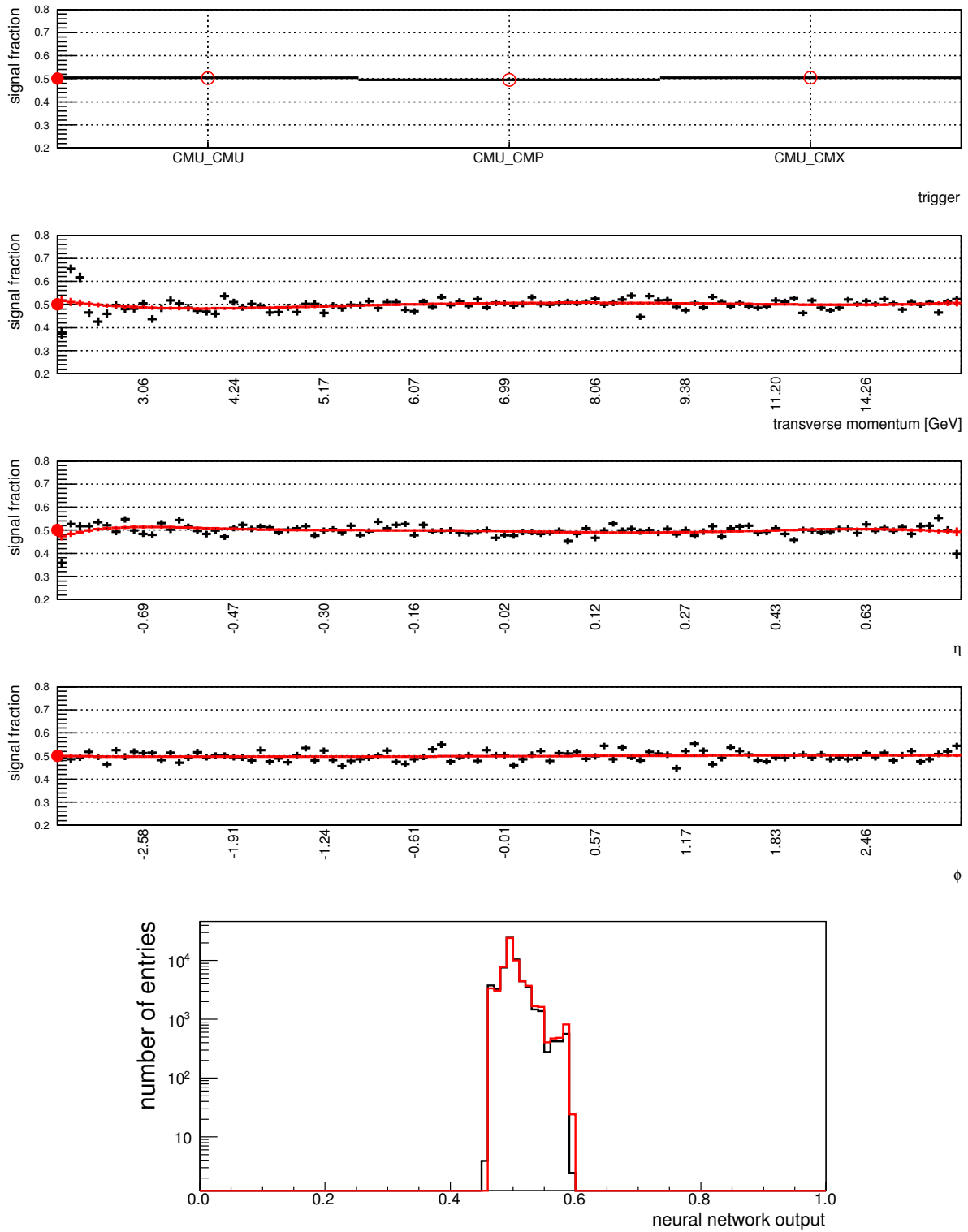
Figure B.2: Validation of the B^+ Monte Carlo reweighting network with the reweighted Monte Carlo sample. The plotting conventions are explained in chapter 4.5.1.

Bibliography

[1] B. Aubert *et al.* (BaBar Collaboration),
"Measurement of B^0-anti-B^0 Flavor Oscillations in Hadronic B^0 Decays",
Phys. Rev. Lett. **88**, 221802 (2002)

[2] B. Aubert *et al.* (BaBar Collaboration),
"Measurement of the B^0-anti-B^0-Oscillation Frequency with Inclusive Dilepton Events",
Phys. Rev. Lett. **88**, 221803 (2002)

[3] T. Tomura *et al.* (Belle Collaboration),
"Measurement of the Oscillation Frequency for B^0-$\bar{B}^0$ Mixing using Hadronic B^0 Decays",
Phys. Lett. B **542**, 207-215 (2002)

[4] K. Hara, M. Hazumi *et al.* (Belle Collaboration),
"Measurement of the B^0-$\bar{B}^0$ Mixing Parameter Δm_d using Semileptonic B^0 Decays",
Phys. Rev. Lett. **89**, 251803 (2002)

[5] A. Abulencia *et al.* (CDF Collaboration),
"Observation of B_s^0-$\bar{B}_s^0$ Oscillations",
Phys. Rev. Lett. **97**, 242003 (2006)

[6] DØ Collaboration,
"Measurement of the Flavor Oscillation Frequency of B_s^0 Mesons",
Conference Note 5474-CONF (2007)

[7] CDF Collaboration,
"An Updated Measurement of the CP Violating Phase $\beta_s^{J/\psi\,\phi}$",
CDF Public Note 08-07-24 (2008)

[8] M. Kobayashi and T. Maskawa,
"CP Violation in the Renormalizable Theory of Weak Interaction",
Prog. Theor. Phys. **49**, 652 (1973)

[9] N. Cabibbo,
"Unitary Symmetry and Leptonic Decays",
Phys. Rev. Lett. **10**, 531 (1963)

[10] The Nobel Prize in Physics 2008, http://nobelprize.org/

[11] L. Wolfenstein,
"Parametrization of the Kobayashi-Maskawa Matrix",
Phys. Rev. Lett. **51**, 1945 (1983)

[12] J. Charles *et al.* (CKMfitter Group),
"CP violation and the CKM matrix: assessing the impact of the asymmetric B factories",
Eur. Phys. J. C **41**, 1-131 (2005)

[13] H. Albrecht *et al.* (ARGUS Collaboration),
Observation of B0 - anti-B0 Mixing,
Phys. Lett. B **192**, 245 (1987)

[14] C. Amsler *et al.* (Particle Data Group),
"Review of Particle Physics",
Phys. Lett. B **667**, 1 (2008)

[15] A. J. Buras, W. Slominski and H. Steger,
"B0 Anti-B0 Mixing, CP Violation And The B Meson Decay",
Nucl. Phys. B **245** 369 (1984)

[16] J. H. Christenson, J. W. Cronin, V. L. Fitch, R. Turlay,
"Evidence for the 2π Decay of the K_2^0 Meson",
Phys. Rev. Lett. **13**, 138 (1964)

[17] A. Lenz,
"Theoretical status of B_s-mixing and lifetimes of heavy hadrons",
Nucl. Phys. Proc. Suppl. **177-178**, 81 (2008)

[18] A. Lenz, U. Nierste,
"Theoretical update of Bs-Bs-bar mixing",
JHEP **0706**, 072 (2007)

[19] H. G. Moser and A. Roussarie,
"Mathematical methods for B0 anti-B0 oscillation analyses",
Nucl. Instrum. Meth. A **384**, 491 (1997)

[20] CDF Collaboration,
"Muon B Flavour Tagging - A Likelihood Approach",
CDF Internal Note 7034 (2004)

[21] CDF Collaboration,
     "Likelihood Based Electron Tagging",
     CDF Internal Note 7121 (2004)

[22] CDF Collaboration,
     "Opposite Side Kaon Tagger",
     CDF Public Note 8241 (2006)

[23] C. Lecci,
     "A Neural Jet Charge Tagger for the Measurement of the B_s^0-$\overline{B}_s^0$ Oscillation Frequency at CDF",
     IEKP-KA/2005-13 (2005)

[24] CDF Collaboration,
     "Combined opposite side flavor tagger",
     CDF Public Note 8460 (2006)

[25] CDF Collaboration,
     "Optimization of the Same Side Kaon tagging algorithm combining PID and kinematic variables",
     CDF Internal Note 8344 (2006)

[26] J. Morlock,
     PhD thesis in preparation

[27] Fermilab Accelerator Devision,
     "Concepts Rookie Book",
     http://www-bdnew.fnal.gov/operations/rookie_books/Concepts_v3.5.pdf

[28] Fermilab Accelerator Devision,
     "Main Injector Rookie Book",
     http://www-bdnew.fnal.gov/operations/rookie_books/Main_Injector_v1.pdf

[29] D. Mohl, G. Petrucci, L. Thorndahl, S. Van Der Meer,
     "Physics and Technique of Stochastic Cooling",
     Phys. Rept. **58**, 73 (1980)

[30] S. Van Der Meer,
     "Stochastic Cooling and the Accumulation of Anti-Protons",
     Rev. Mod. Phys. **57**, 689 (1985)

[31] CDF Collaboration,
     "The CDF II Detector Technical Design Report",
     FERMILAB-Pub-96/390-E (1996)

[32] T. Affolder *et al.*,
     "CDF Central Outer Tracker",
     Nucl. Instrum. Meth. A **526**, 249 (2004)

[33] P. Mack,
     "Kaonindentifikation im Silizium-Vertex-Detektor bei CDF Run II",
     IEKP-KA/2003-24 (2003)

[34] D. Acosta *et al.*,
     "A Time-Of-Flight Detector in CDF-II",
     Nucl. Instrum. Meth. A **518**, 605 (2004)

[35] CDF Collaboration,
     "Prospects for Measuring γ with $B_s^0 \to D_s^{\mp} K^{\pm}$",
     CDF Public Note 4863 (1999)

[36] S. Kuhlmann *et al.*,
     "The CDF calorimeter upgrade for Run IIb",
     Nucl. Instrum. Meth. A **518**, 39 (2004)

[37] L. Balka *et al.*,
     "The CDF Central Electromagnetic Calorimeter",
     Nucl. Instrum. Meth. A **267**, 272 (1988)

[38] S. Bertolucci *et al.*,
     "The CDF Central and Endwall Hadron Calorimeter",
     Nucl. Instrum. Meth. A **267**, 301 (1988)

[39] G. Ascoli *et al.*,
     "CDF Central Muon Detector",
     Nucl. Instrum. Meth. A **268**, 33 (1988)

[40] A. Artikov *et al.*,
     "Design and construction of new central and forward muon counters for CDF II",
     Nucl. Instrum. Meth. A **538**, 358 (2005)

[41] CDF Collaboration,
     "Conceptual Design of a Deadtimeless Trigger for the CDF Trigger Upgrade",
     CDF Public Note 2038 (1994)

[42] E. J. Thomson *et al.*,
     "Online track processor for the CDF upgrade",
     IEEE Trans. Nucl. Sci. **49**, 1063 (2003)

[43] CDF Collaboration,
"Level 2 and the L1/L2 Trigger System for Non-Experts",
CDF Public Note 5860 (2002)

[44] CDF Collaboration,
"CDF – Secondary Vertex Trigger", CDF Public Note 6107 (2002)

[45] M. Feindt,
"A Neural Bayesian Estimator for Conditional Probability Densities",
IEKP-KA/2004-05 (2004)

[46] CDF Collaboration,
"Soft lepton identification using neural networks",
CDF Internal Note 8352 (2006)

[47] T. Allmendinger, G. J. Barker, M. Feindt, C. Haag and M. Moch,
"BSAURUS: A package for inclusive B reconstruction in DELPHI",
arXiv:hep-ex/0102001 (2001)

[48] C. Groß,
"Ein neuronaler Algorithmus zur Unterscheidung von b-Quarks und b-Antiquarks
bei CDF Run II",
IEKP-KA/2005-19 (2005)

[49] A. Schmidt,
"Combined B Flavour Tagging with Artificial Neural Networks",
IEKP-KA/2006-11 (2006)

[50] T. Sjoestrand,
"High–Energy Physics Event Generation with PYTHIA 5.7 and JETSET 7.4.",
Comput. Phys. Commun. **82**, 74 (1994)

[51] CDF Collaboration,
"User Guide For EvtGen @ CDF",
CDF Internal Note 5618 (2003)

[52] F. Abe *et al.* (CDF Collaboration),
"Measurement of the antiproton-proton total cross section at $\sqrt{s} = 546$ and
1800 GeV",
Phys. Rev. D **50**, 5550 (1994)

Danksagung