

KALIBRIERUNG NEUER
'FLAVOR TAGGING' ALGORITHMEN
MITTELS B_s -OSZILLATIONEN

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

genehmigte

Dissertation

von

Dipl. Phys. Philipp Mack
aus Blaubeuren

Tag der mündlichen Prüfung: 13. Juli 2007

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

Korreferent: Prof. Dr. G. Quast, Institut für Experimentelle Kernphysik

Deutsche Zusammenfassung

*„Die Technik von heute ist das Brot von morgen -
die Wissenschaft von heute ist die Technik von morgen.“*

Richard von Weizsäcker

Die Teilchenphysik ist heutzutage eine Wissenschaft, die in großem Maße Elemente aus verschiedenen Fachrichtungen, wie der Mathematik und der Informatik, nutzt, um neue Erkenntnisse über den Ursprung des Universums, seiner Bestandteile und die Interaktion dieser Bestandteile zu gewinnen. Aus einer Vielzahl an Experimenten weiß man, dass die Materie und auch die Antimaterie im Universum aus Elementarteilchen, den Quarks und den Leptonen, aufgebaut ist. Die Quarks können jedoch nicht als freie Teilchen existieren und kommen daher nur als gebundene Zustände, den Hadronen, in der Natur vor. Es gibt zwei Arten von Hadronen, die aus drei Quarks bestehenden Baryonen und die Mesonen, die aus einem Quark und einem Antiquark aufgebaut sind.

Eine besondere Rolle spielen hierbei die neutralen Mesonen K^0 , D^0 , B^0 und B_s , da diese in ihre Antiteilchen oszillieren können. Die Eigenschaften der K^0 - und B^0 -Oszillationen wurden im Laufe der letzten Jahre an vielen Experimenten, vor allem an den B-Fabriken BABAR und BELLE, mit großer Genauigkeit bestimmt [1–3]. B_s -Oszillationen wurden auf Grund ihrer hohen Oszillationsfrequenz jedoch erst im Jahr 2006 am Tevatron beobachtet. Dort wurde am CDF II Experiment die zur Zeit präziseste Messung der Oszillationsfrequenz des B_s Mesons vorgenommen :

$$\Delta m_s = 17.77 \pm 0.10 \pm 0.07 \text{ ps}^{-1} [4].$$

Die zeitaufgelöste Messung der B_s -Oszillationen ist der erste Schritt zur Bestimmung der mischungsinduzierten CP-Verletzung. Diese entsteht durch Interferenzeffekte zwischen Zerfällen von oszillierten und nichtoszillierten B_s -Mesonen in Endzustände, die CP-Eigenzustände sind. Um die CP-Verletzung zu messen, benötigt man neben der Kenntnis der Oszillationsfrequenz auch ein gutes Verständnis über die Leistungsfähigkeit des Algorithmus, der zur Bestimmung des b Flavours zum Produktionszeitpunkt, im folgenden Tagging genannt, verwendet wird. Die Leistungsfähigkeit, im Weiteren mit T abgekürzt, beschreibt die effektive Statistik, die einer Analyse unter Verwendung des Tagging-Algorithmus zur Verfügung steht. Sie ist das Produkt aus der Effizienz ϵ des Algorithmus und der vorhergesagten Dilution D, die ein Maß für die Wahrscheinlichkeit einer richtigen Entscheidung des Algorithmus darstellt.

$$T = \epsilon D^2 \tag{1}$$

Der leistungsstärkste Algorithmus, der momentan am CDF II-Experiment verwendet wird, ist der Same Side Kaon Tagging-Algorithmus. Er erreicht eine Leistungsfähigkeit von ca. 4%.

II

Die Leistungsfähigkeit des Algorithmus wird im Allgemeinen mittels Monte Carlo Simulationen abgeschätzt und muß auf echten Daten verifiziert und kalibriert werden. Im Rahmen dieser Arbeit wurde eine Methode zur Kalibrierung neuer Tagging-Algorithmen mittels der B_s -Oszillation entwickelt und erstmals auf Daten getestet. Um dies zu realisieren wurde zuerst eine komplette Oszillationsanalyse durchgeführt. Für die Bestimmung der Oszillationsfrequenz Δm_s mussten verschiedene Voraussetzungen erfüllt werden. Diese Voraussetzungen können an Hand der Signifikanz des B_s -Oszillationssignals motiviert werden:

$$S = \sqrt{\frac{\epsilon D^2}{2}} e^{-\frac{(\Delta m_s \cdot \sigma_{c\tau})^2}{2}} \frac{N_S}{\sqrt{N_S + N_B}} \quad (2)$$

Dabei stehen N_S und N_B für die Anzahl an Signal-, bzw. Untergrundereignissen der selektierten B_s -Mesonen und $\sigma_{c\tau}$ ist die Auflösung der Lebensdauermessung.

Man benötigt für eine maximal signifikante Messung von Δm_s somit eine optimale Selektion an B_s Mesonen, eine gute Auflösung der Lebensdauer und einen leistungsfähigen Tagging-Algorithmus. Der erste Teil dieser Arbeit befasste sich mit der Optimierung der Selektion exklusiv rekonstruierter B_s -Mesonen. Exklusiv heißt, der Zerfallskanal des B_s -Mesons ist eindeutig bestimmt und alle Teilchen der Zerfallskette werden vollständig rekonstruiert. In dieser Arbeit werden drei verschiedene Zerfallskanäle des B_s -Mesons untersucht:

- $B_s \rightarrow D_s \pi$ und $D_s \rightarrow K^* K$
- $B_s \rightarrow D_s 3\pi$ und $D_s \rightarrow K^* K$
- $B_s \rightarrow D_s 3\pi$ und $D_s \rightarrow 3\pi$

Des Weiteren wurden die Ergebnisse für die Optimierung der folgenden Zerfallskanäle, die alle in der Karlsruher B-Physik Gruppe studiert wurden [5, 6], in die Analyse integriert:

- $B_s \rightarrow D_s \pi$ und $D_s \rightarrow \phi \pi$
- $B_s \rightarrow D_s \pi$ und $D_s \rightarrow 3\pi$
- $B_s \rightarrow D_s 3\pi$ und $D_s \rightarrow \phi \pi$

Um die große Anzahl an Parametern und Funktionen, die für die Analyse dieser Kanäle und die Oszillationsanalyse erforderlich sind, einfach handhaben zu können, wurde eine modulare Programmstruktur entwickelt. Diese Programmstruktur ist nicht nur die Grundlage für diese Analyse, sondern wird inzwischen auch in der B-Meson-Spektroskopie [7] und für die Bestimmung der Lebensdauerdifferenz $\Delta\Gamma$ im B_s -System eingesetzt.

Die Optimierung der Selektion von B_s -Mesonen wurde mit Hilfe von neuronalen Netzwerken durchgeführt. Diese besitzen im Gegensatz zur herkömmlichen schnittbasierten Optimierung den Vorteil einer komplexen Selektion im n -dimensionalen Phasenraum der Variablen und berücksichtigten somit die Korrelationen der Variablen, welche für die Trennung von Signal und Untergrund verwendet werden. Für die Auswertung der Leistungsfähigkeit der neuronalen Netze wird die Signifikanz der B_s -Selektion $\frac{N_S}{\sqrt{N_S + N_B}}$ durch Fits an die Massenverteilungen der Zerfallskanäle bestimmt. In dieser Analyse wurden ungefähr 4600 B_s Signalereignisse aus einem Datensatz mit einer integrierten Luminosität von ca. 1 fb^{-1} rekonstruiert.

Dabei konnte die Signifikanz der B_s -Selektion je nach betrachtetem Zerfallskanal um 15-90% im Vergleich zur schnittbasierten Methode gesteigert werden.

Die Lebensdauerermessung des B_s -Mesons ist eine weitere notwendige Zutat zur Bestimmung der Oszillationsfrequenz. Dazu wurde im Anschluß an die Optimierung der B_s -Meson-Selektion die Lebensdauer des B_s -Mesons, der Massenmittelwert und die Signalbreite des B_s -Mesons mittels eines ungebinnten Maximum-Likelihood-Fits gemessen. Die Ergebnisse dieses Fits stimmen im Rahmen der Unsicherheiten mit den Weltmittelwerten überein.

Die bei der Lebensdauerermessung bestimmten Ergebnisse bilden die Grundlage für die Δm_s -Messung. Diese wird ebenfalls mittels eines ungebinnten Maximum-Likelihood-Fits bestimmt. Die zugehörige Likelihood Funktion kann wie folgt ausgedrückt werden :

$$L = \prod_i (1 - f_b) \cdot \frac{1}{N_S} \left(\frac{1 + \xi_i D_i \cos(\Delta m_s t_i)}{1 + |\xi_i|} \cdot e^{-\frac{t_i}{\tau}} \right) \otimes \mathcal{G}(t_i - t', \sigma_{t,i}) \cdot \epsilon(t_i) \cdot P_S(m_i) + f_b \cdot (1 + \xi_i D_{bg}) \cdot P_B(t_i) \cdot P_B(m_i) \quad (3)$$

Hier steht f_b für den Anteil an Untergrundereignissen und N_S ist die Normierung des Signalanteils. Die Größe ξ gibt für jedes Ereignis die Vorhersage an, ob das B_s -Meson oszilliert hat oder nicht. Sie wird aus der Entscheidung des Tagging-Algorithmus und der Ladung des B_s -Mesons zum Zerfallszeitpunkt bestimmt. Der Signalanteil der Likelihoodfunktion besteht aus dem Oszillationsterm, der durch eine Faltung mit der gauß'schen Auflösungsfunktion der Lebensdauerermessung verschmiert ist, und der limitierten Akzeptanz, die durch die Effizienzfunktion $\epsilon(t)$ beschrieben wird. Der Untergrundanteil der Likelihoodfunktion besteht aus den Wahrscheinlichkeitsdichten der Massen- und Lebensdauererzeugungen und der Korrektur auf mögliche Asymmetrien des Tagging-Algorithmus. Dabei ist D_{bg} die Untergrund-Dilution, auch sie wird im Maximum-Likelihood Fit bestimmt. Für das Tagging wurde neben dem Same Side Kaon Tagging-Algorithmus ein neuer auf neuronalen Netzwerken basierender Tagging-Algorithmus im Maximum-Likelihood-Fit eingesetzt. Dieser Tagging-Algorithmus wurde in der Karlsruher B-Physik Gruppe entwickelt und besitzt eine Effizienz von 100%, d.h. für jedes Ereignis wird eine Tagging-Entscheidung geliefert.

Die Ergebnisse der ungebinnten Maximum-Likelihood-Fits für die Kombination der optimierten Selektion der B_s -Mesonen aus den Zerfallskanälen $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$, $K^* K$, 3π sind

$$\Delta m_s = 17.89_{-0.14}^{+0.13} \text{ ps}^{-1} \quad \text{für den Same Side Kaon Tagging-Algorithmus}$$

$$\Delta m_s = 17.86 \pm 0.12 \text{ ps}^{-1} \quad \text{für den Karlsruhe Neuronale Netzwerk Tagging-Algorithmus}$$

Die Zerfallskanäle $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi \pi$, $K^* K$, 3π wurden für den kombinierten Fit nicht berücksichtigt, da die Tagging-Information für den Karlsruhe Neuronale Netzwerk Tagging-Algorithmus auf den untersuchten Daten nicht vorhanden ist. Die hier präsentierten Ergebnisse für die Oszillationsfrequenz sind im Rahmen der statistischen Unsicherheiten kompatibel mit der in [4] gemessenen Oszillationsfrequenz von $\Delta m_s = 17.77 \pm 0.10 \pm 0.07 \text{ ps}^{-1}$.

IV

Da im Rahmen dieser Analyse eine Asymmetrie im Zerfallskanal $B_s \rightarrow D_s$, $D_s \rightarrow \phi\pi$ für B_s - und \bar{B}_s -Mesonen zum Zerfallszeitpunkt festgestellt wurde und die gemessenen Oszillationsfrequenzen in diesem Kanal für die unterschiedlichen Tagging-Algorithmen um 0.8 ps^{-1} abweichen, wurden zwei unterschiedliche Szenarien für die anschließende Kalibrierung untersucht. Das erste Szenario besteht aus der Kombination aller untersuchten Zerfallskanäle $B_s \rightarrow D_s\pi$. Im zweiten Szenario wurde nur die Kombination der Zerfallskanäle $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$ und $B_s \rightarrow D_s\pi$, $D_s \rightarrow 3\pi$ untersucht.

Zur Kalibrierung des Tagging-Algorithmus wird die Methode des Amplitudenscans verwendet. Dazu führt man in den Oszillationsterm der Likelihood-Funktion eine zusätzliche Amplitude A ein :

$$\frac{1 + \xi_i D_i \cos(\Delta m_s t_i)}{1 + |\xi_i|} \rightarrow \frac{1 + \xi_i A D_i \cos(\Delta m_s t_i)}{1 + |\xi_i|}$$

Im Gegensatz zum normalen Amplitudenscan wird in dieser Arbeit die Amplitude nur für die im ungebinnten Maximum-Likelihood Fit gemessene Oszillationsfrequenz bestimmt, anstatt über viele Oszillationsfrequenzen zu scannen. Die Amplitude ist nun ein Maß für die wirkliche Leistungsfähigkeit des Tagging-Algorithmus. Sie skaliert die vorhergesagte Dilution. Die wirkliche Leistungsfähigkeit des Tagging-Algorithmus ergibt sich zu :

$$T = \epsilon D^2 = \epsilon \frac{1}{n} \sum_{i=1}^n (A D_i)^2$$

n ist die Anzahl der Ereignisse der B_s -Meson-Selektion.

Aus den gemessenen Amplituden konnten nun die kalibrierten Leistungsfähigkeiten für die unterschiedlichen Szenarien bestimmt werden. Man erhält folgende Ergebnisse :

Szenario 1 : $B_s \rightarrow D_s\pi$ und $D_s \rightarrow \phi\pi, K^*K, 3\pi$

$T = 4.4 \pm 3.0\%$ für den Same Side Kaon Tagging-Algorithmus

$T = 4.8 \pm 3.1\%$ für den Karlsruher Neuronalen Netzwerk Tagging-Algorithmus

Szenario 2 : $B_s \rightarrow D_s\pi$ und $D_s \rightarrow K^*K, 3\pi$

$T = 5.7 \pm 5.1\%$ für den Same Side Kaon Tagging-Algorithmus

$T = 7.5 \pm 5.7\%$ für den Karlsruher Neuronalen Netzwerk Tagging-Algorithmus

Die angegebenen Unsicherheiten sind nur statistischer Natur, da systematische Unsicherheiten auf Grund der Dominanz der statistischen Unsicherheiten vernachlässigt werden können. Die Ergebnisse für den Karlsruher Neuronalen Netzwerk Tagging-Algorithmus sind bereits jetzt vielversprechend, und die Kalibration soll in Kürze auf einem Datensatz mit einer integrierten Luminosität von ca. 2 fb^{-1} wiederholt werden. Außerdem werden die $B_s \rightarrow D_s 3\pi$ -Zerfallskanäle in die bisherige Analyse integriert, sobald die Tagging-Information für den Karlsruher Neuronalen Netzwerk Tagging-Algorithmus für diese Kanäle auf den Daten erhältlich ist.

Des Weiteren muß das Problem der Asymmetrie im Zerfallskanal $B_s \rightarrow D_s, D_s \rightarrow \phi\pi$ verstanden werden, um die Kalibrierung der Tagging-Algorithmen weiter zu verbessern. Erreicht dann der Karlsruher Neuronale Netzwerk Tagging-Algorithmus seine vorhergesagte Leistungsfähigkeit von 9%, bedeutet dies eine Verdopplung gegenüber der Leistungsfähigkeit bisheriger Tagging-Algorithmen.

Der erste Test der Kalibrierung von Tagging-Algorithmen mittels B_s -Oszillationen zeigt, dass die in dieser Arbeit entwickelte Methode funktioniert und in Zukunft für die Messung der mischungsinduzierten CP-Verletzung verwendet werden kann.

CALIBRATION OF NEW FLAVOR TAGGING ALGORITHMS
USING B_s OSCILLATIONS

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

genehmigte

Dissertation

von

Dipl. Phys. Philipp Mack
aus Blaubeuren

Tag der mündlichen Prüfung: 13. Juli 2007

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

Korreferent: Prof. Dr. G. Quast, Institut für Experimentelle Kernphysik

Contents

Introduction	11
1 Theoretical Overview	13
1.1 The Standard Model	13
1.2 The CKM Matrix	14
1.3 Meson Oscillations	17
1.3.1 Oscillation Phenomenology	17
1.3.2 Standard model predictions for B_q^0 oscillation	19
1.3.3 Status of oscillation measurements	20
2 The CDF II Experiment	23
2.1 The Accelerator	23
2.1.1 The Accelerator Complex	23
2.1.2 Accelerator performance	27
2.2 The CDF Detector	29
2.2.1 Tracking System	30
2.2.2 Particle Identification	31
2.2.3 Calorimeters	32
2.2.4 Muon Detector System	33
2.3 The Run II Trigger System	34
2.3.1 Level 1 Trigger	34
2.3.2 Level 2 Trigger	34
2.3.3 Level 3 Trigger	35
3 Statistical Methods	37
3.1 Parameter Estimation	37
3.1.1 The Maximum Likelihood Method	37
3.2 Artificial neural networks	41
3.2.1 The Feed Forward Neural Network	43
3.2.2 Training of a neural network	43
3.2.3 Interpretation of the network output	44
3.2.4 NeuroBayes [®]	44
4 Analysis Outline	47
4.1 Overview of the Analysis	47
4.1.1 Mixing probability	48
4.1.2 Significance of the Δm_s Measurement	49

4.2	Data Selection and Event Reconstruction	49
4.2.1	The Two Track Trigger	50
4.2.2	Event Reconstruction	50
4.3	Monte Carlo Simulation	52
4.4	The Fitter Framework	54
5	B Flavor Tagging	57
5.1	General Properties of B Flavor Tagging	57
5.1.1	Dilution and Efficiency	57
5.1.2	Opposite Side Tagging	58
5.1.3	Same Side Tagging	59
5.2	B Flavor Tagging at CDF II	59
5.2.1	Karlsruhe Neural Network B Flavor Tagger	61
6	Exclusive B_s Mesons Selection	65
6.1	Motivation	65
6.2	Optimizing the Selection of Exclusive B_s Mesons	65
6.2.1	The Mass Spectra of the B_s Decay Channels	66
6.2.2	Training and Application of Neural Networks in the Selection of Exclusively Reconstructed B_s Mesons	68
6.2.3	Fitting the Mass Spectra	75
7	Lifetime Fit	83
7.1	Proper Decay Time	83
7.1.1	Scale Factor Determination for the Proper Decay Time Resolution	84
7.2	The Lifetime Fit	87
8	Mixing Fit	93
8.1	The Mixing Likelihood	93
8.2	The Amplitude Scan Method	99
9	Conclusion and Outlook	103
A	List of Variables	105
B	Compilation of Training Results	109
C	Compilation of Fit Templates	117
C.1	Fit Templates in Mass Space	117
C.1.1	Cabibbo suppressed signal	117
C.1.2	$D_s X$ background	118
C.1.3	Λ_b background	119
C.1.4	B^0 background	120
C.2	Fit Templates in Lifetime Space	121
C.2.1	Efficiency Function	121
C.2.2	Background Contributions to the Proper Decay Time	122

D	Compilation of Fit Results	125
D.1	Binned Fits	125
D.1.1	Mass Fits	125
D.2	Unbinned Fits	128
D.2.1	Projections of the Fit Results to the Mass and Lifetime Space	128
	Bibliography	137

List of Figures

1.1	Unitarity triangle	16
1.2	Constraints on the unitarity triangle	17
1.3	Box diagrams contributing to the $B_q^0 - \bar{B}_q^0$ mixing	19
1.4	K^0/\bar{K}^0 mixing	21
1.5	D^0/\bar{D}^0 mixing	21
1.6	B^0/\bar{B}^0 mixing	22
1.7	B_s^0/\bar{B}_s^0 mixing	22
2.1	Schematic view of the Fermilab accelerator chain.	24
2.2	The Cockcroft-Walton accelerator.	24
2.3	The LINAC.	24
2.4	The Main Injector and Recycler	25
2.5	CDF II integrated Luminosity	28
2.6	CDF II peak Luminosity.	28
2.7	Cutaway view of the CDF II detector.	29
2.8	Cutaway view of the inner part of the CDF II detector. The acronyms are explained in the corresponding paragraphs of this chapter.	30
2.9	r-z view of the tracking system	32
2.10	The TOF and COT dE/dx separation power	33
2.11	Universal curve of the COT dE/dx for positive tracks	33
2.12	View of the dataflow of the CDF II Trigger System	35
2.13	Diagram of the Level 1 and Level 2 Trigger of CDF II Trigger System	35
2.14	SVT impact parameter resolution	36
2.15	Definition of the track impact parameter d_0 for a SVT track	36
3.1	Sketch of the anatomy of a neuron	42
3.2	Layout of an artificial neuron	42
3.3	Examples of the sigmoid functions for different values of c	42
3.4	Schematic view of a three layer feed forward network	42
3.5	Example of a purity-efficiency plot to determine a working point.	45
3.6	The signal purity plotted against the output of the neural network.	45
4.1	Road-map of the Δm_s measurement.	48
4.2	Leading order diagrams for heavy-flavor production.	53
4.3	Pair creation diagrams	53
4.4	Gluon splitting diagrams	53
4.5	Overview of the components of the fitter framework.	55

4.6	Illustration of the transfer of parameters between the Mode and the Fitter . . .	55
5.1	Schematic view of a B_s event	57
5.2	The cascade decay $b \rightarrow c \rightarrow s$	58
5.3	Fragmentation processes of the same side b quark	59
5.4	CDF II event display of a $b\bar{b}$ pair	60
5.5	Workflow of the flavor tagger on track level.	62
5.6	Flowchart of the flavor tagger on event level.	63
6.1	Example of a typical mass spectrum for the decay of a B_s meson.	66
6.2	Tree level diagrams for the $B_s \rightarrow D_s(3)\pi$ decay	67
6.3	Classification procedure using NeuroBayes [®]	69
6.4	Example of a data and and a MC sample for network training	71
6.5	Helicity angle of the kaon for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	73
6.6	Invariant mass of the K^* for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	73
6.7	Invariant mass of the 3π vertex for the decay $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$	73
6.8	Distribution of $\theta(\vec{n}_{(\pi_{B_s}^1, \pi_{B_s}^2)}, \vec{p}(3\pi))$ for the decay $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$	73
6.9	Invariant mass distribution of $m_{\pi_{D_s}^2, \pi_{D_s}^3}$ for the decay $B_s \rightarrow D_s3\pi, D_s \rightarrow 3\pi$	73
6.10	Binary tree of the general mass fit function	76
6.11	Significance versus network output for $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	77
6.12	Invariant mass spectrum (4.7,6.0) for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	78
6.13	Pull distribution of the mass for 5000 toy experiments with 3000 events	81
6.14	Pull distribution of the width for 5000 toy experiments with 3000 events	82
7.1	Scale factor distributions for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	86
7.2	Scale factor distributions for the decay $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$	86
7.3	Efficiency function for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	88
7.4	Fit of P_{Λ_b} to the Λ_b MC template for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	89
7.5	Fit of P_{B^0} to the B^0 MC template for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	89
7.6	Fit of P_{comb} to the upper sideband for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	89
7.7	Projection of the fit result to the lifetime space for $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	91
7.8	Projection of the fit result to the mass space for $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	91
8.1	Likelihood ratio of a single event	95
8.2	Likelihood ratio for the $B_s \rightarrow D_s^- \pi^+, D_s^- \rightarrow \phi \pi^-$ decay mode	95
8.3	Likelihood ratio for the $\bar{B}_s \rightarrow D_s^+ \pi^-, D_s^+ \rightarrow \phi \pi^+$ decay mode	95
8.4	Graphical representation of the results obtained by the mixing fit.	96
8.5	Likelihood ratio for the $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi$ decay mode	97
8.6	Likelihood ratio for the $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$ decay mode	97
8.7	Likelihood ratio for the combination of the $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K, 3\pi$ decay modes	97
8.8	Likelihood ratio for the combination of the $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi, K^*K, 3\pi$ decay modes	97
8.9	Pull distribution of Δm_s for 5000 toy experiments with 3000 events	98
8.10	Pull distribution of Δm_s for 5000 toy experiments with 30000 events	98
8.11	Squared predicted dilution of the SSK and KANN tagging algorithms for signal like events of the $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$ decay in the narrow mass range.	101

8.12	Amplitude scan for the combined fit using the KANNT	101
8.13	Amplitude scan for the combined fit using the SSKT	102
A.1	Definition of the lts. impact parameters	107
A.2	Definition of the angle $\theta(\vec{n}_{(\pi_{B_s}^1, \pi_{B_s}^2)}, \vec{p}(3\pi))$	107
B.1	Correlation matrix and significance table for $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi$	110
B.2	Quality plots from the training for $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi$	110
B.3	Correlation matrix and significance table for $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	111
B.4	Quality plots from the training for $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	111
B.5	Correlation matrix and significance table for $B_s \rightarrow D_s\pi, D_s \rightarrow \pi\pi\pi$	112
B.6	Quality plots from the training for $B_s \rightarrow D_s\pi, D_s \rightarrow \pi\pi\pi$	112
B.7	Correlation matrix and significance table for $B_s \rightarrow D_s3\pi, D_s \rightarrow \phi\pi$	113
B.8	Quality plots from the training for $B_s \rightarrow D_s3\pi, D_s \rightarrow \phi\pi$	113
B.9	Correlation matrix and significance table for $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$	114
B.10	Quality plots from the training for $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$	114
B.11	Correlation matrix and significance table for $B_s \rightarrow D_s3\pi, D_s \rightarrow \pi\pi\pi$	115
B.12	Quality plots from the training for $B_s \rightarrow D_s3\pi, D_s \rightarrow \pi\pi\pi$	115
C.1	MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi$	117
C.2	MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	117
C.3	MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi$	117
C.4	MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s3\pi, D_s \rightarrow \phi\pi$	117
C.5	MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$	118
C.6	MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s3\pi, D_s \rightarrow 3\pi$	118
C.7	MC fit template for the D_sX background of the $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi$ decay	118
C.8	MC fit template for the D_sX background of the $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$ decay	118
C.9	MC fit template for the D_sX background of the $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi$ decay	118
C.10	MC fit template for the D_sX background of the $B_s \rightarrow D_s3\pi, D_s \rightarrow \phi\pi$ decay	118
C.11	MC fit template for the D_sX background of the $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$ decay	119
C.12	MC fit template for the D_sX background of the $B_s \rightarrow D_s3\pi, D_s \rightarrow 3\pi$ decay	119
C.13	MC fit template for the Λ_b contribution of the $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$ decay	119
C.14	MC fit template for the Λ_b contribution of the $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi$ decay	119
C.15	MC fit template for the Λ_b contribution of the $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$ decay	119
C.16	MC fit template for the Λ_b contribution of the $B_s \rightarrow D_s3\pi, D_s \rightarrow 3\pi$ decay	119
C.17	MC fit template for the B^0 contribution of the $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	120
C.18	MC fit template for the B^0 contribution of the $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$ decay	120
C.19	MC fit template for the B^0 contribution of the $B_s \rightarrow D_s3\pi, D_s \rightarrow 3\pi$ decay	120
C.20	Parameterized fit to the efficiency for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi$	121
C.21	Parameterized fit to the efficiency for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	121
C.22	Parameterized fit to the efficiency for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi$	121
C.23	Parameterized fit to the efficiency for the decay $B_s \rightarrow D_s3\pi, D_s \rightarrow \phi\pi$	121
C.24	Parameterized fit to the efficiency for the decay $B_s \rightarrow D_s3\pi, D_s \rightarrow K^*K$	122
C.25	Parameterized fit to the efficiency for the decay $B_s \rightarrow D_s3\pi, D_s \rightarrow 3\pi$	122
C.26	Fit to the ct of the upper sideband for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi$	122
C.27	Fit to the ct of the upper sideband for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	122
C.28	Fit to the ct of the upper sideband for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi$	123

C.29	Fit to the ct of the upper sideband for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow \phi\pi$. . .	123
C.30	Fit to the ct of the upper sideband for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow K^*K$. .	123
C.31	Fit to the ct of the upper sideband for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow 3\pi$. . .	123
C.32	Fit to the ct of a Λ_b MC sample for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	123
C.33	Fit to the ct of a Λ_b MC sample for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi$	123
C.34	Fit to the ct of a Λ_b MC sample for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow K^*K$	124
C.35	Fit to the ct of a Λ_b MC sample for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow 3\pi$	124
C.36	Fit to the ct of a B^0 MC sample for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$	124
C.37	Fit to the ct of a B^0 MC sample for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow K^*K$. . .	124
C.38	Fit to the ct of a B^0 MC sample for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow 3\pi$	124
D.1	Inv. mass spectrum for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi, \text{nnout} > 0.68$	125
D.2	Inv. mass spectrum for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi, \text{cut based}$	125
D.3	Inv. mass spectrum for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K, \text{nnout} > 0.92$. . .	126
D.4	Inv. mass spectrum for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K, \text{cut based}$	126
D.5	Inv. mass spectrum for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi, \text{nnout} > 0.48$	126
D.6	Inv. mass spectrum for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi, \text{cut based}$	126
D.7	Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow \phi\pi, \text{nnout} > 0.90$.	127
D.8	Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow \phi\pi, \text{cut based}$. . .	127
D.9	Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow K^*K, \text{nnout} > 0.99$	127
D.10	Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow K^*K, \text{cut based}$. .	127
D.11	Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow 3\pi, \text{nnout} > 0.99$. .	128
D.12	Projection of the fit result to the lifetime space for $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi$. . .	128
D.13	Projection of the fit result to the mass space for $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi$	128
D.14	Projection of the fit result to the lifetime space for $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$. .	129
D.15	Projection of the fit result to the mass space for $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K$. . .	129
D.16	Projection of the fit result to the lifetime space for $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi$. . .	129
D.17	Projection of the fit result to the mass space for $B_s \rightarrow D_s\pi, D_s \rightarrow 3\pi$	129
D.18	Projection of the fit result to the lifetime space for $B_s \rightarrow D_s 3\pi, D_s \rightarrow \phi\pi$. .	130
D.19	Projection of the fit result to the mass space for $B_s \rightarrow D_s 3\pi, D_s \rightarrow \phi\pi$	130
D.20	Projection of the fit result to the lifetime space for $B_s \rightarrow D_s 3\pi, D_s \rightarrow K^*K$.	130
D.21	Projection of the fit result to the mass space for $B_s \rightarrow D_s 3\pi, D_s \rightarrow K^*K$. .	130
D.22	Projection of the fit result to the lifetime space for $B_s \rightarrow D_s 3\pi, D_s \rightarrow 3\pi$. .	131
D.23	Projection of the fit result to the mass space for $B_s \rightarrow D_s 3\pi, D_s \rightarrow 3\pi$	131

List of Tables

1.1	Overview of the properties of the elementary fermions	14
1.2	Summary of the properties of the gauge bosons	14
1.3	Current parameters of the four neutral oscillating meson pairs	20
2.1	Accelerator parameters for the Tevatron Run II	27
2.2	Design parameters of the muon chambers	33
4.1	Summary of the datasets used in this analysis.	49
6.1	Summary of the hadronic decay modes	65
6.2	Summary of preselection cuts	70
6.3	Upper sideband mass ranges	70
6.4	Variables used in the neural network trainings	74
6.5	Results of the binned fits for the $B_s \rightarrow D_s\pi$ modes in the wide mass range . .	79
6.6	Results of the binned fits to the $B_s \rightarrow D_s3\pi$ modes in the wide mass range .	79
6.7	Overview of the cut based analysis	80
6.8	Recalculated fractions of the $B_s \rightarrow D_s\pi$ decay modes in the narrow mass range.	80
6.9	Recalculated fractions of the $B_s \rightarrow D_s3\pi$ decay modes in the narrow mass range.	80
6.10	Results of the toy MC experiments in mass space for $B_s \rightarrow D_s\pi$	81
7.1	Parameters p for $D_s\pi$, xbhd0d datasets	85
7.2	Parameters p for $D_s\pi$, xbhd0h+i datasets	85
7.3	Parameters p for $D_s3\pi$, xbhd0d datasets	85
7.4	Parameters p for $D_s3\pi$, xbhd0h+i datasets	85
7.5	Results of the unbinned lifetime fits for the $B_s \rightarrow D_s\pi$ modes.	90
7.6	Results of the unbinned lifetime fits for the $B_s \rightarrow D_s3\pi$ modes.	92
7.7	Results of the toy MC experiments in the lifetime space	92
8.1	Results of the mixing fit.	96
8.2	Results of the toy MC experiments for the unbinned mixing fit	98
8.3	The fitted amplitudes and the corresponding tagging powers.	100

Introduction

The field of elementary particle physics is one of the cornerstones for comprehending the basic principles the universe is based on. In elementary particle physics recent experiments use gigantic accelerators and detectors to resolve distances as small as a billionth of the size of an atom and to reproduce the events happening nanoseconds after the big bang. Our current knowledge of the basic components of the universe and their interactions is described by the standard model of particle physics. Matter and antimatter in our universe consist of elementary particles, the quarks and the leptons. Quarks do not exist as free particles but as bound quark states, called hadrons. There are two different kinds of hadrons, the baryons consisting of three quarks and the mesons composed of a quark anti-quark pair. A special kind of mesons are the neutral mesons K^0 , D^0 , B^0 and B_s . They are the only mesons which are able to oscillate into their own antiparticle and back. The oscillation frequency corresponds to the mass difference of the two mass eigenstates of the particle. The mass eigenstates of the meson resemble those of a system of two coupled pendula of the same length. The solution where the pendula oscillate in phase corresponds to a light mass eigenstate and the solution where the pendula oscillates with a phase difference of 180° corresponds to a heavy mass eigenstate. Today the properties of the meson oscillation have been measured precisely in the K^0/\bar{K}^0 and B^0/\bar{B}^0 system [1–3]. In 2006 B_s oscillations have first been observed at the Tevatron. The CDF II experiment provides at present the best measurement of the oscillation frequency $\Delta m_s = 17.77 \pm 0.10 \pm 0.07 \text{ ps}^{-1}$ [4]. And most recently the first evidence for D^0 - \bar{D}^0 oscillations has been seen at the B factories at the BABAR [8] and BELLE [9] experiments.

The time-resolved measurement of the B_s oscillation is the starting point of many interesting analyses. The next important effect one wants to study after the measurement of the B_s oscillation is the mixing induced CP violation in the B_s/\bar{B}_s system. The mixing induced CP violation has its origin in the interference of amplitudes of the mixed and unmixed decay to final states common to B_s and \bar{B}_s . To measure the mixing induced CP violation one needs a flavor tagged time dependent analysis. Therefore the flavor of the b quark at production time is determined by a flavor tagging algorithm. The recent flavour tagging algorithm used at the CDF II experiment has a tagging power of about four percent [10]. The tagging power is the effective statistics corresponding to the number of events with a perfect tag. Many of the current analyses are hindered by the low tagging power. Therefore a new flavor tagging algorithm was developed in our physics group combining all information used in previous tagging algorithms with the help of artificial neural networks. The new flavor tagging algorithm has been tested on simulated Monte Carlo samples and a tagging power of about nine percent has been estimated [11].

The main goal of this thesis is to prepare the B_s mixing analysis for the calibration of the tagging algorithms. Calibrating the tagging algorithm means the determination of the true tagging power on data by measuring the mixing frequency and the corresponding amplitude of the time dependent asymmetry at this frequency. It is the first time the calibration of tagging algorithms is attempted in a B_s mixing analysis. Furthermore it is the first test to estimate the potential of the new flavor tagging algorithm on data.

The structure of the thesis reflects the necessary course of action for the mixing analysis and the successive calibration of the tagging algorithm. The first chapter gives a summary of the theory of B_s oscillations. In the second chapter the accelerator complex at the Tevatron and the CDF II experiment are described. The third chapter points out the statistical methods used throughout the thesis focusing on the maximum likelihood method and the principles of artificial neural networks. The fourth chapter explains the acquisition of the data and Monte Carlo samples and presents the fitter framework used for the mixing analysis. In the fifth chapter the physics of flavor tagging, the tagging algorithms currently used at the CDF II experiment and the new Karlsruhe neural network tagging algorithm are described. The sixth chapter deals with the optimization of the exclusive selection of B_s mesons. Hence six different decay channels of the B_s meson, namely the decays $B_s \rightarrow D_s(3)\pi$ with $D_s \rightarrow \phi\pi, K^*K, 3\pi$, are studied. In order to provide an optimal selection of B_s mesons to the final mixing fit, artificial neural networks are used instead of a cut based approach. Furthermore the probability density functions used for the part of the final maximum likelihood function concerning the mass space are introduced. The parameters for these functions are then determined by binned mass fits. After the selection of the B_s meson candidates and the mass fits, the description of the probability density function of the lifetime space is the next necessary ingredient for the mixing analysis. The characterization of the lifetime space and the measurement of the proper decay time of the B_s meson in the studied decay channels are subject to chapter seven. Finally the ingredients needed for the mixing analysis as presented in the previous chapters are put together in chapter eight. Thereupon the mixing frequency Δm_s is measured by performing an unbinned maximum likelihood fit. Afterwards the tagging algorithms are calibrated at the measured mixing frequency. In chapter nine the results for the mixing frequency and the calibrated tagging algorithms are summarized and an outlook is given.

Chapter 1

Theoretical Overview

1.1 The Standard Model

The standard model of elementary particle physics [12–15] is up to now the best theory to comprehend the properties of particles and their interactions. Its mathematical description underlies the local symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. The Lagrangian is invariant under gauge transformations in this symmetry. In the standard model the interactions of particles are mediated by the gauge bosons, spin-one particles. The particles with spin 1/2 are called fermions. The physical gauge bosons can be expressed in terms of the gauge fields of the symmetry groups :

$$SU(3) \rightarrow G_\mu^{\alpha=1..8}, SU(2) \rightarrow W_\mu^{\alpha=1..3} \text{ and } U(1) \rightarrow B_\mu$$

- the gluon, which mediates the strong force :
 $G_\mu^\alpha, \alpha = 1, \dots, 8$
- the W boson, the charged gauge boson of the weak interaction :
 $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \pm iW_\mu^2)$
- the Z boson, the neutral gauge boson of the weak interaction :
 $Z_\mu^0 = \cos(\theta_W)W_\mu^3 - \sin(\theta_W)B_\mu$
- the photon, responsible for the electromagnetic interaction :
 $A_\mu = \sin(\theta_W)W_\mu^3 + \cos(\theta_W)B_\mu$

$$\text{with the weak mixing angle } \sin(\theta_W) = 0.23122(15) \quad [16]$$

The fermions can be divided into quarks and leptons. The leptons only couple to the weak and if charged to the electromagnetic force. The quarks however possess an additional quantum number, the color charge and therefore participate in strong interactions. Tables 1.1 and 1.2 summarize basic properties of fermions and gauge bosons^{1 2}.

¹In this thesis, energy, momentum and mass are given in GeV. 1 GeV = $1.602176462(63) \cdot 10^{-10}$ J. The units kg m/s and kg for momentum and mass can be obtained by the division by c , respectively c^2 , where $c = 299792458$ m/s is the speed of light.

²In the standard model the neutrino masses are set to zero. However various experiments show that their masses are not zero [17–21] and therefore the standard model has to be extended [22].

Quarks			Leptons		
flavor	m [GeV]	q [e]	flavor	m [MeV]	q [e]
up u	$(1.5 - 3.0) \times 10^{-3}$	+2/3	electron e	$0.51099892 \pm (4 \times 10^{-8})$	-1
down d	$(3.0 - 7.0) \times 10^{-3}$	-1/3	el. neutrino ν_e	$< 4.6 \times 10^{-4}$	0
charm c	1.25 ± 0.09	+2/3	muon μ	105.658369 ± 0.000009	-1
strange s	$(95 \pm 25) \times 10^{-3}$	-1/3	muon neutrino ν_μ	< 0.19	0
top t	174.2 ± 3.3	+2/3	tau τ	$1776.99^{+0.29}_{-0.26}$	-1
bottom b	4.20 ± 0.07	-1/3	tau neutrino ν_τ	< 18.2	0

Table 1.1: Overview of the properties of the elementary fermions [16]

Name	m [GeV]	q[e]	Interaction
Photon γ	0	0	electromagnetic
W^\pm	80.403 ± 0.029	± 1	weak
Z^0	91.1876 ± 0.0021	0	weak
Gluon	0	0	strong

Table 1.2: Summary of the properties of the gauge bosons [16]

In order to generate the mass of quarks, charged leptons and the gauge bosons without loosing the local gauge invariance of the Lagrangian, one introduces the Higgs mechanism. It is described by a scalar field ϕ with a non-vanishing vacuum expectation value in the electroweak standard model. This field is realized by the Higgs boson and the mechanism of spontaneous symmetry breaking then generates the masses of the particles mentioned above [23].

1.2 The CKM Matrix

The different families of fermions are connected via the charged W^\pm gauge bosons. This charged current is the only mechanism in the Standard Model which allows for flavor changes and can be written as [24]:

$$J_\mu^{cc} = \begin{pmatrix} \bar{\nu}_e \\ \bar{\nu}_\mu \\ \bar{\nu}_\tau \end{pmatrix} \gamma_\mu \frac{1 - \gamma_5}{2} \begin{pmatrix} e \\ \mu \\ \tau \end{pmatrix} + \sum_{r,g,b} \begin{pmatrix} \bar{u} \\ \bar{c} \\ \bar{t} \end{pmatrix} \gamma_\mu \frac{1 - \gamma_5}{2} \mathbf{V}_{\text{CKM}} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

where \mathbf{V}_{CKM} is the 3×3 transformation matrix between quark mass eigenstates and flavor eigenstates, called Cabibbo-Kobayashi-Maskawa matrix. In its most general form it can be expressed by [25, 26]:

$$\mathbf{V}_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

As the elements V_{ij} of the CKM matrix are complex, they generate 18 free real parameters. If there are only three quark families the CKM matrix is a general unitary matrix with the

requirement $\mathbf{V}_{\text{CKM}}\mathbf{V}_{\text{CKM}}^\dagger = \mathbf{1}$. This reduces the number of free parameters to 9. Of the 9 parameters, 5 can be absorbed in one global phase and 4 relative phases between the u,c,t and d,s,b quarks. Those phases are all subject to convention and have no physical meaning. Thus the CKM matrix can be described by 4 real parameters, one phase parameter and three rotation angles in flavor space. The standard parameterization, noted below, uses phases that leave V_{ud} and V_{cd} real [27] :

$$\mathbf{V}_{\text{CKM}} = \begin{pmatrix} c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta_{13}} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta_{13}} & c_{13}c_{23} \end{pmatrix}$$

with $c_{ij} = \cos \theta_{ij}$ and $s_{ij} = \sin \theta_{ij}$, where θ_{ij} are the rotation angles and δ_{13} is the complex phase. An angle of historical importance is $\theta_C = \theta_{12}$, the Cabibbo angle, that describes the mixing between the first two generations.

A more popular parameterization of the CKM matrix is the Wolfenstein parameterization, which uses the parameters [28] :

- $\lambda = \sin \theta_C$
- $\eta = \sin \theta_{13} \sin \delta_{13} / A\lambda^3$
- $\rho = \sin \theta_{13} \cos \delta_{13} / A\lambda^3$
- $A = \sin \theta_{23} / \lambda^2$

An expansion of λ up to the third order leads to the following description :

$$\mathbf{V}_{\text{CKM}} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4)$$

The unitarity of the CKM matrix requires the following conditions for the matrix elements to be fulfilled :

$$\sum_{i=1}^3 \mathbf{V}_{ij} \mathbf{V}_{ik}^* = \delta_{jk} = \sum_{i=1}^3 \mathbf{V}_{ji} \mathbf{V}_{ki}^* \quad j=1..3, k=1..3$$

The most frequently used unitarity condition is the product of the first and third column of the CKM matrix :

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \quad (1.1)$$

Dividing equation 1.1 by $A\lambda^3 \approx -V_{cd}V_{cb}^*$ and using the Wolfenstein approximation yields :

$$(\rho + i\eta) - 1 + (1 - \rho - i\eta) = 0$$

This corresponds to the graphical representation of the unitarity triangle displayed in figure 1.1. The coordinates of the corners of the triangle are $(0, 0)$, $(1, 0)$ and $(\bar{\rho}, \bar{\eta})$, with

$$\bar{\rho} = \left(1 - \frac{\lambda^2}{2}\right)\rho \quad , \quad \bar{\eta} = \left(1 - \frac{\lambda^2}{2}\right)\eta$$

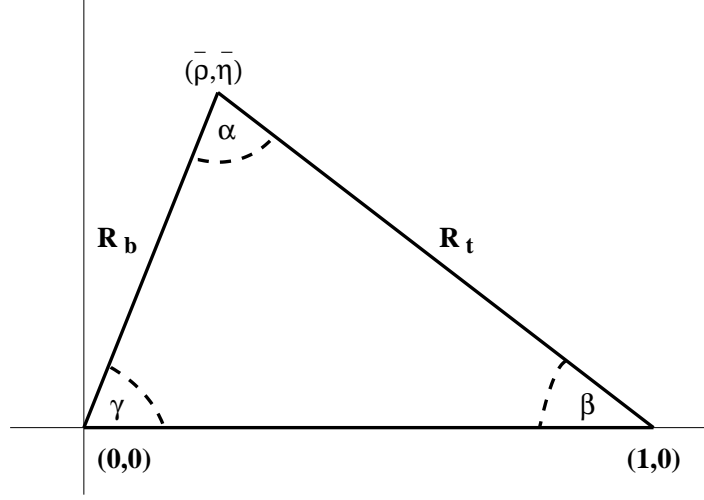


Figure 1.1: Unitarity triangle in the commonly used Wolfenstein representation

The sides of the triangle, apart the side along the real axis which is normalized to unity, are specified by :

$$R_b = \left| \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} \right| = \sqrt{\bar{\rho}^2 + \bar{\eta}^2}$$

$$R_t = \left| \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right| = \sqrt{(1 - \bar{\rho})^2 + \bar{\eta}^2}$$

The angles can be defined as [16] :

$$\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right)$$

Because the CKM matrix elements are fundamental parameters of the standard model, their precise determination is an important task in the field of particle physics. This is achieved by overconstraining the unitary triangle through independent experimental measurements of the sides and angles. Using these measurements one is able to validate the standard model and constrain the impact of physics beyond the standard model to flavor changing interactions. In figure 1.2 the current status of the experimental constraints on the parameters of the unitarity triangle is shown. The values for the Wolfenstein parameters and the CKM matrix elements, derived by a global fit to the experimental data, are [29, 30] :

$$\mathbf{V}_{\text{CKM}} = \begin{pmatrix} 0.97383_{-0.00023}^{+0.00024} & 0.2272_{-0.0010}^{+0.0010} & (3.96_{-0.09}^{+0.09}) \times 10^{-3} \\ 0.2271_{-0.0010}^{+0.0010} & 0.97296_{-0.00024}^{+0.00024} & (42.21_{-0.80}^{+0.10}) \times 10^{-3} \\ (8.14_{-0.64}^{+0.32}) \times 10^{-3} & (41.61_{-0.78}^{+0.12}) \times 10^{-3} & 0.999100_{-0.000004}^{+0.000034} \end{pmatrix}$$

$$\lambda = 0.2272 \pm 0.0010 \quad A = 0.818_{-0.017}^{+0.007}$$

$$\bar{\rho} = 0.221_{-0.028}^{+0.064} \quad \bar{\eta} = 0.340_{-0.045}^{+0.017}$$

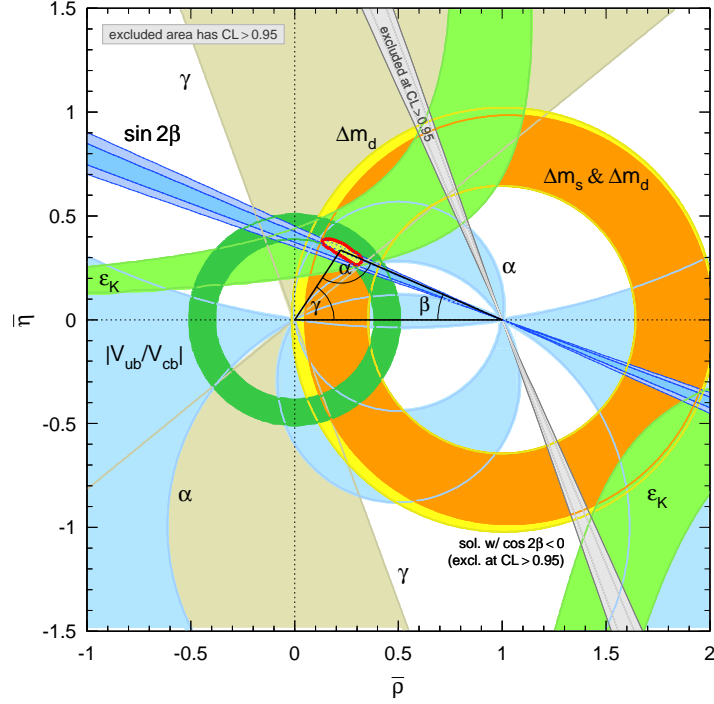


Figure 1.2: Constraints on the parameters of the unitarity triangle in the $\bar{\rho}$ - $\bar{\eta}$ plane before the measurement of Δm_s [16].

1.3 Meson Oscillations

Mesons consist of a colored quark and an antiquark of the corresponding anti-color, e.g. red and anti-red. Therefore the meson is colorless. A special class of mesons are pairs of neutral charge-conjugate mesons, which can oscillate into each other via flavor changing weak interaction transitions. These are $K^0/\bar{K}^0(\bar{s}d/s\bar{d})$, $D^0/\bar{D}^0(c\bar{u}/\bar{c}u)$, $B^0/\bar{B}^0(\bar{b}d/b\bar{d})$ and $B_s^0/\bar{B}_s^0(\bar{b}s/b\bar{s})$. The following section explains the oscillation phenomenology of the B_s in detail, but it will also be valid as a general description for the other meson oscillations.

1.3.1 Oscillation Phenomenology

The non-relativistic Schrödinger equation for a charge-conjugate meson pair is :

$$i\partial_t\psi = \mathbf{H}\psi$$

with a general non-hermitian Hamiltonian [31]

$$\mathbf{H} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \begin{pmatrix} m_{11} - \frac{i}{2}\Gamma_{11} & m_{12} - \frac{i}{2}\Gamma_{12} \\ m_{12}^* - \frac{i}{2}\Gamma_{12}^* & m_{22} - \frac{i}{2}\Gamma_{22} \end{pmatrix}$$

CPT invariance requires $m_{11} = m_{22} = m$ and $\Gamma_{11} = \Gamma_{22} = \Gamma$.

The indices 1 and 2 of the Hamiltonian represent the orthogonal and normalized base vectors :

$$|B_s^0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |\bar{B}_s^0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The solutions of the Schrödinger equation for meson pairs are the same as for the mechanical analogon of two coupled pendula of the same length and correspond to a short-lived, heavy eigenstate $B_{s,H}$ and a long-lived, light eigenstate $B_{s,L}$. These mass eigenstates can be expressed in terms of the flavor eigenstates by the equations

$$|B_{s,L}\rangle = p|B_s^0\rangle + q|\bar{B}_s^0\rangle$$

$$|B_{s,H}\rangle = p|B_s^0\rangle - q|\bar{B}_s^0\rangle$$

with the normalization $|p|^2 + |q|^2 = 1$ and the ratio η_m defined as :

$$\eta_m = \frac{q}{p} = -2 \frac{m_{12}^* - i\frac{\Gamma}{2}\Gamma_{12}^*}{m_H - m_L - \frac{i}{2}(\Gamma_H - \Gamma_L)}$$

Introducing the dimensionless parameters x, y and the scaled time variable T

$$x = \frac{m_H - m_L}{\Gamma} = \frac{\Delta m}{\Gamma}, \quad y = \frac{\Gamma_H - \Gamma_L}{2\Gamma} = \frac{\Delta\Gamma}{2\Gamma}$$

$$\Gamma = \frac{1}{2}(\Gamma_H + \Gamma_L), \quad T = \Gamma t$$

the time evolution of an arbitrary initial state

$$|\psi(0)\rangle = a|B_s^0\rangle + \bar{a}|\bar{B}_s^0\rangle$$

can be written as [32]

$$\begin{aligned} |\psi(t)\rangle &= e^{-imt-T/2} \left[(a|B_s^0\rangle + \bar{a}|\bar{B}_s^0\rangle) \cos\left((x-iy)\frac{T}{2}\right) \right. \\ &\quad \left. + i\left(\frac{\bar{a}}{\eta_m}|B_s^0\rangle + a\eta_m|\bar{B}_s^0\rangle\right) \sin\left((x-iy)\frac{T}{2}\right) \right] \end{aligned} \quad (1.2)$$

The experimental observable is the decay rate to a flavor specific final state X and \bar{X} at a given time T . It can be obtained by multiplying $|\psi(t)\rangle$ with $\langle X|\mathbf{H}$, respectively $\langle \bar{X}|\mathbf{H}$ and then integrating the amplitudes over the phase space :

$$\dot{N}_{B_s^0 \rightarrow X}(t) = N_0 \int d\text{PS} |\langle X|\mathbf{H}|\psi(t, \bar{a}=0)\rangle|^2 = \frac{1}{2} N_0 e^{-T} \Gamma_X (\cosh yT + \cos xT) \quad (1.3)$$

$$\dot{N}_{B_s^0 \rightarrow \bar{X}}(t) = N_0 \int d\text{PS} |\langle \bar{X}|\mathbf{H}|\psi(t, \bar{a}=0)\rangle|^2 = \frac{1}{2} N_0 |\eta_m|^2 e^{-T} \Gamma_X (\cosh yT - \cos xT) \quad (1.4)$$

with the partial width for a non oscillating meson

$$\Gamma_X = \int d\text{PS} |\langle X|\mathbf{H}|B_s^0\rangle|^2 = \int d\text{PS} |\langle X|\mathbf{H}|\bar{B}_s^0\rangle|^2$$

Using the equations above, the asymmetry, showing the difference of the numbers of mesons and anti-mesons as a function of the scaling lifetime, can be expressed via

$$a(T) = \frac{\dot{N}_{X \rightarrow X} - \dot{N}_{X \rightarrow \bar{X}}}{\dot{N}_{X \rightarrow X} + \dot{N}_{X \rightarrow \bar{X}}} \Big|_T = \frac{(1 - |\eta_m|^2) \cosh yT + (1 + |\eta_m|^2) \cos xT}{(1 + |\eta_m|^2) \cosh yT + (1 - |\eta_m|^2) \cos xT} \quad (1.5)$$

This equation is fundamental to the determination of the mixing frequency. The measurement of the time-dependent asymmetry is used to extract Δm and is explained in chapter 8 of this thesis.

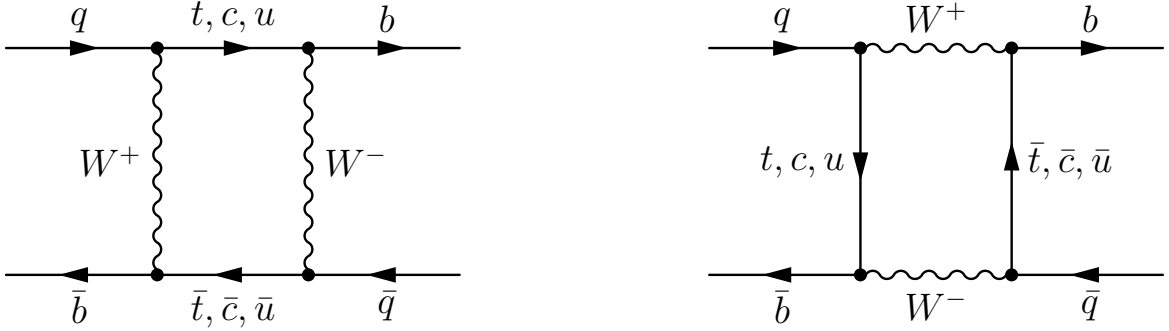


Figure 1.3: Box diagrams contributing to the $B_q^0 - \bar{B}_q^0$ mixing. The label q represents either a d (B^0) or a s (B_s^0) quark

1.3.2 Standard model predictions for B_q^0 oscillation

The Hamiltonian \mathbf{H} can be divided into a strong, electromagnetic part \mathbf{H}_0 and a weak interaction perturbation \mathbf{H}_w . In the Wigner-Weisskopf approximation [33] the leading term for $\mathbf{H}_{12,21}$ is $\langle B_q^0 | \mathbf{H} | \bar{B}_q^0 \rangle$, which corresponds to the box diagrams shown in 1.3. An approximate evaluation of these diagrams using the Feynman rules yields to [34]:

$$m_{12} = -\frac{G_F^2}{12\pi^2} e^{i\phi_{CP}} V_{tb}^2 V_{tq}^{*2} m_W^2 m_{B_q} [f_{B_q}^2 B_q] \cdot [S(m_t^2/m_W^2) \cdot \eta_{QCD}] \quad (1.6)$$

$$\Delta m = 2|m_{12}| = \frac{G_F^2}{6\pi^2} |V_{tb}^* V_{tq}|^2 m_W^2 m_{B_q} [f_{B_q}^2 B_q] \cdot [S(m_t^2/m_W^2) \cdot \eta_{QCD}] \quad (1.7)$$

$$\Delta\Gamma \approx -m_{12} \cdot \frac{3\pi}{2S(m_t^2/m_W^2)} \frac{m_b^2}{m_W^2} \left[1 + \frac{8}{3} \frac{m_c^2}{m_b^2} \frac{V_{cb} V_{cq}^*}{V_{tb} V_{tq}^*} + \mathcal{O}\left(\frac{m_c^4}{m_b^4}\right) \right] \quad (1.8)$$

Here $S(m_t^2/m_W^2)$ denotes the Inami-Lim function [35]

$$S(x) = x \left[\frac{1}{4} + \frac{9}{4(1-x)} - \frac{3}{2(1-x)^2} - \frac{3x^2 \ln x}{2(1-x)^3} \right]$$

η_{QCD} is a perturbative QCD correction to the Inami-Lim function. In a consistent renormalization scheme the values are $S \approx 2.3$ and $\eta_{QCD} \approx 0.55$ [36]. f_{B_q} is the B decay constant and B_q is the vacuum insertion correction. The largest theoretical uncertainty is introduced by $f_{B_q}^2 B_q$. At present the best result for this value is calculated from lattice gauge theory.

The ratio $\Delta\Gamma/\Delta m$ can be estimated with the relation $\Delta m \cdot \Delta\Gamma = 4 \operatorname{Re}(m_{12}\Gamma_{12}^*)$ [32] and the equations 1.7 and 1.8 to be

$$\frac{\Delta\Gamma}{\Delta m} = \frac{2y}{x} \approx -\frac{3\pi}{2} \frac{m_b^2}{m_t^2} \approx \frac{1}{250}$$

Expressing $|V_{tb}^*V_{td}|$ in the Wolfenstein parameterization leads to

$$|V_{tb}^*V_{td}| = \sqrt{\lambda^6 A^2 [(1 - \bar{\rho})^2 + \bar{\eta}^2] + \mathcal{O}(\lambda^{10})}$$

Inserting this in equation 1.7 shows that Δm_d describes a circle around $(1,0)$ in the $\bar{\rho} - \bar{\eta}$ plane. For a better constraint on the unitarity triangle, one can use the ratio

$$\frac{\Delta m_s}{\Delta m_d} = \frac{m_{B_s^0} f_{B_s^0}^2 B_s^0}{m_{B^0} f_{B^0}^2 B^0} \left| \frac{V_{ts}}{V_{td}} \right|^2 = \frac{m_{B_s^0}}{m_{B^0}} \xi^2 \left| \frac{V_{ts}}{V_{td}} \right|^2$$

In this ratio the theoretical uncertainties nearly cancel out, leaving only contributions from ξ^2 , which can be extracted from lattice QCD with a precision around 5% to a value of $\xi = 1.21_{-0.035}^{+0.047}$ [37]. With the recent measurement of $\Delta m_s = 17.77 \pm 0.10 \pm 0.07 \text{ ps}^{-1}$ [4] of the CDF II collaboration and the world average values of $\Delta m_d = 0.507 \pm 0.005 \text{ ps}^{-1}$ [38], $m(B_d)/m(B_s) = 0.98390$ [39], one obtains $|V_{ts}/V_{td}| = 0.2060 \pm 0.0007(\text{stat}+\text{sys})_{-0.0060}^{+0.0081}(\text{theory})$ [4].

1.3.3 Status of oscillation measurements

Recently the current status of the measurements of meson oscillations have changed drastically, especially in the D and B meson system. The newest results come from the B factories, BABAR and BELLE, which have seen first evidence for $D^0-\bar{D}^0$ oscillations [8,9]. In table 1.3 the latest measurements of the parameters of the meson systems, respectively the theoretical expectations if no measurements exists, are summarized.

The figures 1.4 to 1.7 display the normalized number of mesons (anti-mesons) and the time dependent asymmetry plotted against the scaled time variable T . These figures clearly point out the differences in the meson systems, e.g. the very fast oscillating B_s^0 -system or the very slow oscillating D^0 -system, where the oscillation is hardly visible.

	K^0/\bar{K}^0	D^0/\bar{D}^0	B^0/\bar{B}^0	B_s^0/\bar{B}_s^0
$\tau[\text{ps}]$	89.53 ± 0.05 51140 ± 210	0.4101 ± 0.0015	1.530 ± 0.009	1.466 ± 0.059
$\Gamma[\text{ps}^{-1}]$	5.584×10^{-3}	2.438	0.654	0.682
$ y $	0.9984	0.0131 ± 0.0032 [9]	$< 0.01^*$ [40]	$\approx 0.1^*$ [41]
$\Delta m[\text{ps}^{-1}]$	$(5.292 \pm 0.009) \times 10^{-3}$	< 0.12	0.507 ± 0.005	17.77 ± 0.10 [4]
x	0.9477	< 0.05	0.7757	26.05
$ \eta_m ^2$	0.99348 ± 0.00024	$\approx 1^*$ [40]	$1 \dots 1.002^*$ [40]	$\approx 1^*$ [40]

Table 1.3: Current parameters of the four neutral oscillating meson pairs. Values are taken from [16], unless specified. Standard model predictions are indicated with an asterisk.

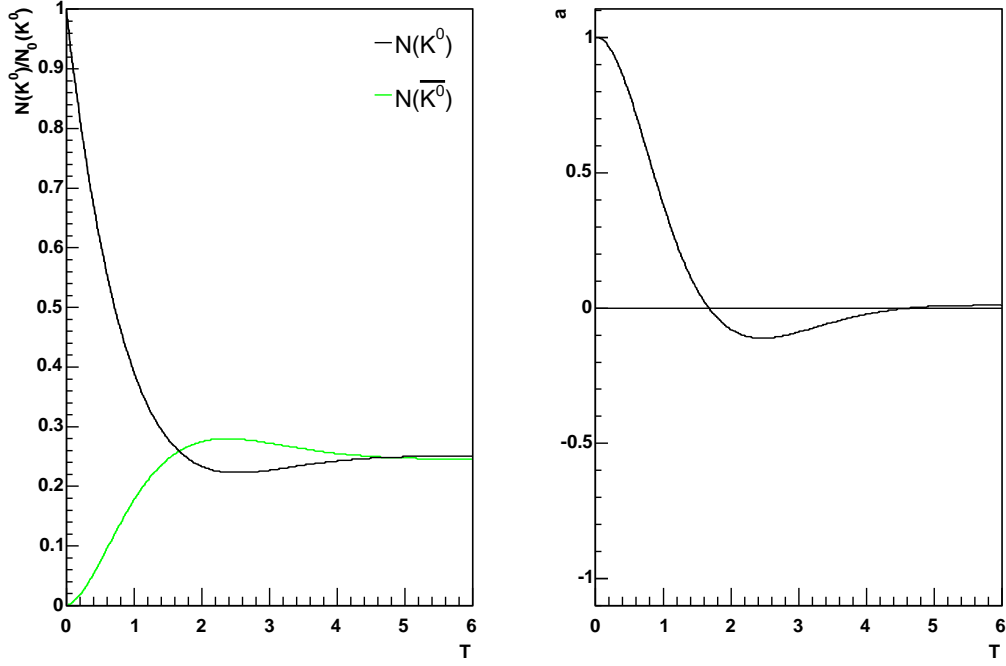


Figure 1.4: K^0/\bar{K}^0 mixing. The number of K^0 and \bar{K}^0 as function of T (left). The asymmetry $a = (N_{K^0} - N_{\bar{K}^0})/(N_{K^0} + N_{\bar{K}^0})$ (right). Values of $x = 0.9477$, $y = -0.9984$ and $|\eta_m|^2 = 0.99348$ were used for the creation of the diagrams.

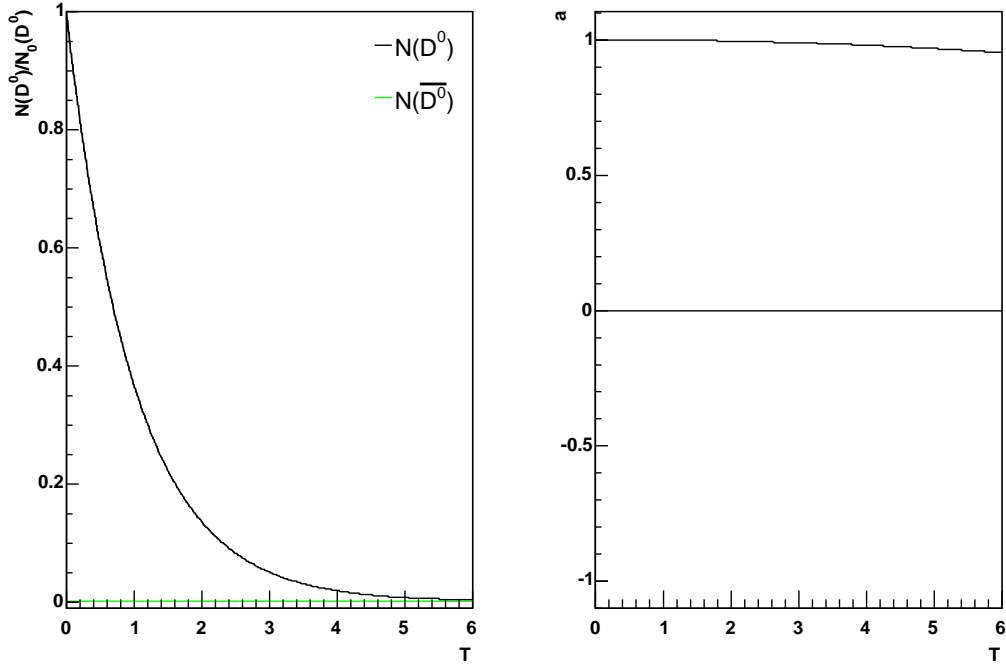


Figure 1.5: D^0/\bar{D}^0 mixing. The number of D^0 and \bar{D}^0 as function of T (left). The asymmetry $a = (N_{D^0} - N_{\bar{D}^0})/(N_{D^0} + N_{\bar{D}^0})$ (right). These plots were generated with a large $x = 0.05$, $y = 0.013$ and $|\eta_m|^2 = 1$.

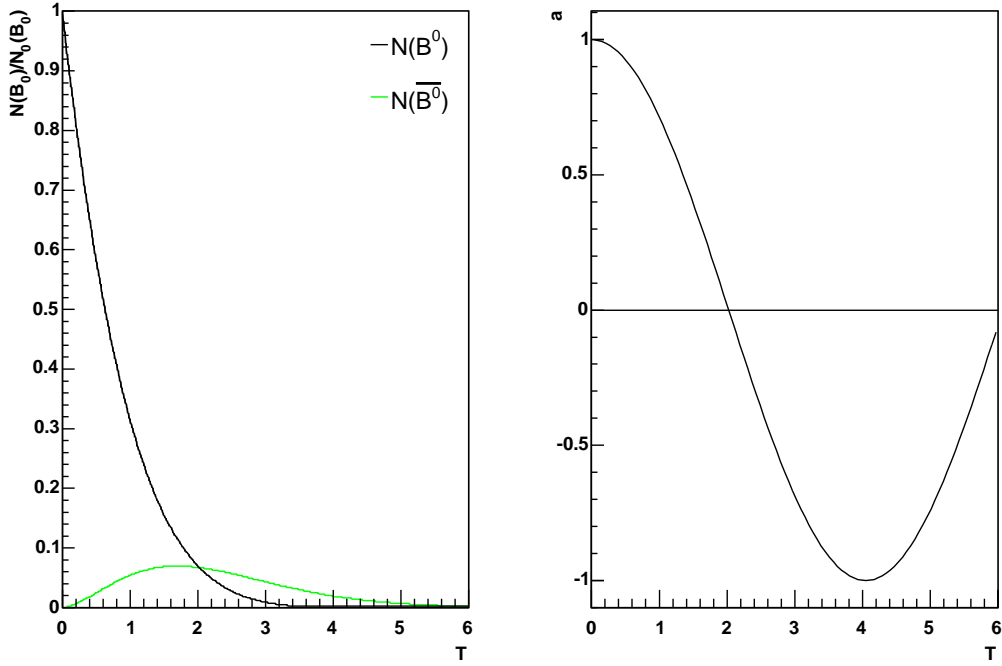


Figure 1.6: B^0/\bar{B}^0 mixing. The number of B^0 and \bar{B}^0 as function of T (left). The asymmetry $a = (N_{B^0} - N_{\bar{B}^0})/(N_{B^0} + N_{\bar{B}^0})$ (right). The plots correspond to the parameters $x = 0.7757$, $y = 0$ and $|\eta_m|^2 = 1$.

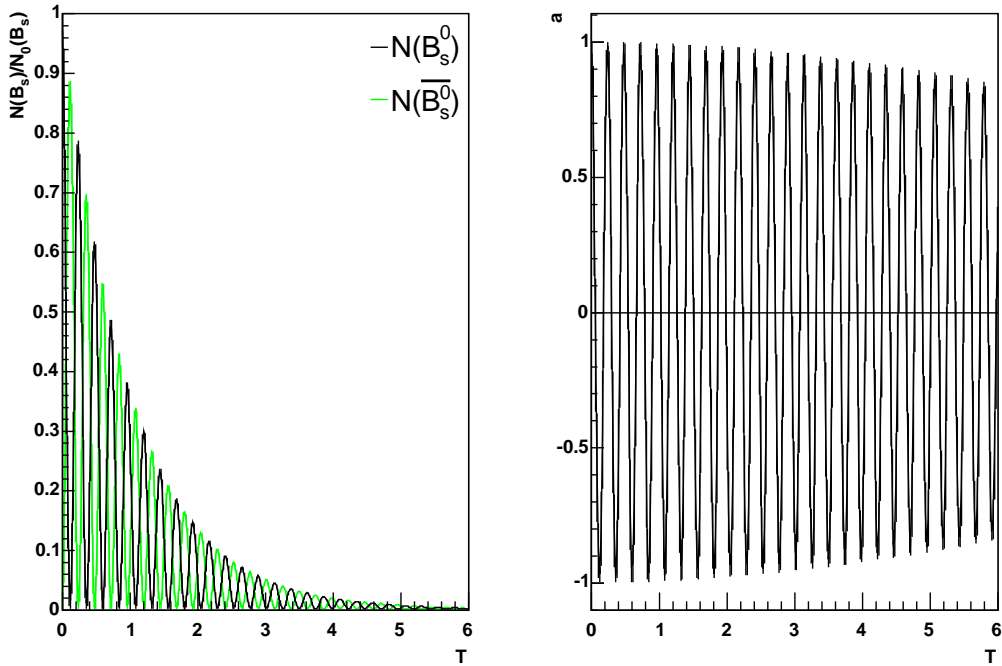


Figure 1.7: B_s^0/\bar{B}_s^0 mixing. The number of B_s^0 and \bar{B}_s^0 as function of T (left). The asymmetry $a = (N_{B_s^0} - N_{\bar{B}_s^0})/(N_{B_s^0} + N_{\bar{B}_s^0})$ (right). These plots relate to values of $x = 26$, $y = 0.01$ and $|\eta_m|^2 = 1$.

Chapter 2

The CDF II Experiment

The Fermi National Accelerator Laboratory (Fermilab) is situated about 70 km west of Chicago in Batavia/Illinois. It is host to the Tevatron, a proton-antiproton ($p\bar{p}$) collider, which operates at a center of mass energy of 1.96 TeV, currently the highest energy for $p\bar{p}$ collisions. At the interaction regions of the proton and antiproton beams the two multipurpose detectors DØ and CDF (Collider Detector at Fermilab) are located.

2.1 The Accelerator

2.1.1 The Accelerator Complex

To achieve the acceleration of protons and antiprotons to a beam energy of 980 GeV, a chain of accelerators is used. The Fermilab accelerator complex is shown in figure 2.1 and consists of the following parts [42] :

- Cockcroft-Walton pre-accelerator :

In the Cockcroft-Walton pre-accelerator negatively charged hydrogen ions are produced by surface ionization effects. Those ions are then accelerated to an energy of 750 keV and injected into the linear accelerator. The device is shown in figure 2.2 .

- LINear ACcelerator (LINAC) :

The LINAC is a 150 m long linear accelerator. It is made of cavities in which the hydrogen ions are accelerated by oscillating radio frequency (RF) fields up to an energy of 400 MeV. At the end of the LINAC the hydrogen ions are grouped into bunches automatically because of the geometry of the RF fields. The hydrogen ions are then transferred to the Booster. A picture of the LINAC is presented in figure 2.3 .

- Booster :

The Booster is a circular accelerator with a radius of 75 m. At the entrance of the booster the electrons of the hydrogen ions are stripped off while passing a carbon foil, so that only bare protons remain. The protons are then accelerated to a kinetic energy of 8 GeV. This is achieved by a sequence of kicks from RF cavities.

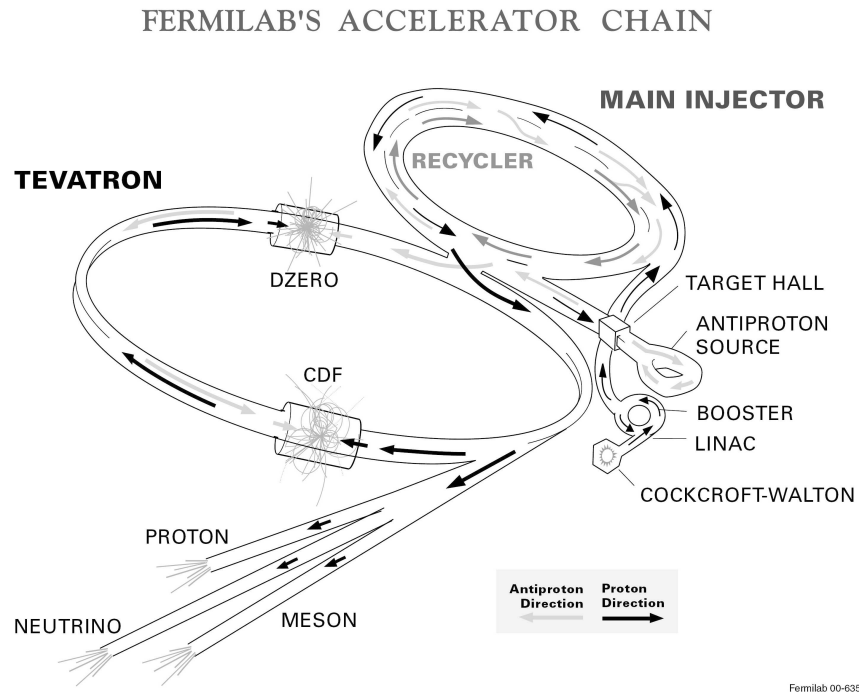


Figure 2.1: Schematic view of the Fermilab accelerator chain.

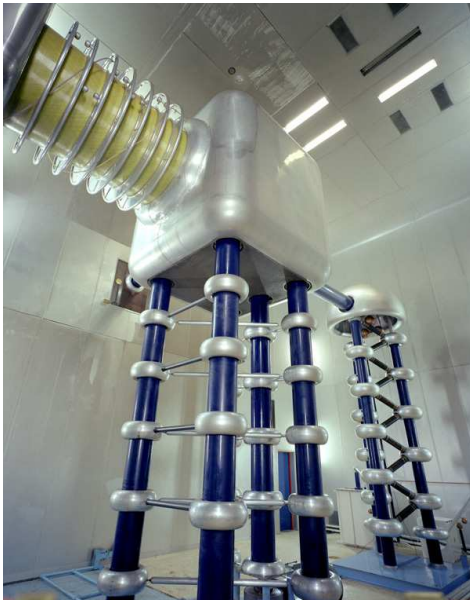


Figure 2.2: The Cockcroft-Walton accelerator.



Figure 2.3: The LINAC.



Figure 2.4: The Main Injector (light blue magnets at the bottom) and the Recycler (green magnets on top).

In each revolution around the Booster the protons gain about 500 keV in kinetic energy. Additional protons are injected to increase the intensity of the proton beam. Then the beam is passed on to the Main Injector.

- Main Injector :

The Main Injector, a 3 km circumference synchrotron is used for several different functions. It accelerates the protons to an energy of 150 GeV. Secondly it provides protons with an energy of 120 GeV for the antiproton production. In addition it receives the antiprotons from the antiproton source and accelerates them to an energy of 150 GeV. Subsequently it injects protons and antiprotons into the Tevatron. It also provides particle beams to several fixed target and neutrino experiments. A part of the Main Injector ring is displayed in figure 2.4 .

- Antiproton Source :

To create antiprotons the 120 GeV proton beam from the main injector is directed onto a nickel target. During the collision of the protons with the nickel atoms many secondary particles are created, among those the antiprotons. The antiprotons are then focused into a beam with a Lithium lens. Afterwards the antiprotons are filtered from the other particles by sending the beam through a pulsed magnet. Because the proton beam from the Main Injector is bunched, the antiprotons coming off the target are also bunched and will have a large spread in energy.

To reduce this energy spread and to debunch the antiproton beam, the Debuncher accelerator uses different RF phases in the cavities. After the Debuncher the antiprotons are piled up in the Accumulator Synchrotron. This is achieved by momentum stacking successive pulses of antiprotons from the Debuncher over several hours or days. The Accumulator Synchrotron uses both RF and stochastic cooling systems for this process. If enough antiprotons are accumulated, they are directed to the Main Injector.

- The Tevatron

The final stage of the accelerator chain is the Tevatron synchrotron. Its ring has a radius of 1 km and consists of approximately 1000 superconducting magnets, each of them operating at a temperature of 4.3 Kelvin and generating a magnetic field of 4.2 Tesla. During the operation of the Tevatron 36 bunches of protons and antiprotons are simultaneously in the ring and are accelerated to the final center of mass energy of 1.96 TeV. Each bunch contains about 3×10^{11} protons respectively 9×10^{10} antiprotons. After the beams reach their full energy, they are focused further and brought to collision in interaction regions, where the two experiments, CDF II and DØ are located. Each collision reduces the number of particles in the beam so that after a certain period of time, called store, protons and antiprotons have to be reinjected from the Main Injector. The duration of a store is usually about one day.

- The Antiproton Recycler

The Antiproton Recycler was the latest addition to the accelerator chain in the year 2004. It was placed in the Main Injector ring and operates at a fixed energy of 8 GeV. It functions as a post-Accumulator ring by receiving antiprotons from the Accumulator if the stacking rate decreases. Thus it allows the Accumulator to operate at an optimum antiproton intensity regime. Because it is built of permanent magnets, the recycler has the advantage of surviving even a power glitch and thus keeping the stash of stored antiprotons. Figure 2.4 shows a part of the Antiproton Recycler.

2.1.2 Accelerator performance

The quantity in which one measures the performance of an accelerator is the luminosity. The interaction probability of two particles with each other is described by the cross-section σ_{int} . The interaction rate within the detector can be written as :

$$\dot{N} = \sigma_{int}\mathcal{L}$$

where \mathcal{L} is the luminosity. The luminosity defines how the particles in the proton and anti-proton bunches interact and can be expressed as follows [43] :

$$\mathcal{L} = F\left(\frac{\sigma_l}{\beta^*}\right)\frac{fN_bN_pN_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)}$$

where f is the revolution frequency, N_p and $N_{\bar{p}}$ are the number of protons, respectively anti-protons, in the bunch, N_b is the number of bunches, σ_p^2 and $\sigma_{\bar{p}}^2$ are the standard deviations of the beam size at the interaction point, $F\left(\frac{\sigma_l}{\beta^*}\right)$ is a form factor, dependent on the bunch length σ_l , and the beta function β^* at the interaction point. In table 2.1 the values for the parameters of the accelerator in Run II are listed. The performance of an experiment can be determined by the integration of the luminosity over the time of a data taking period. To estimate the number of events of a given process, one multiplies the integrated luminosity with the cross section σ of this process :

$$N = \sigma \times \int \mathcal{L}$$

Until the year 2007 a peak luminosity of about $2.8 \times 10^{32} cm^{-2}s^{-1}$ has been achieved and an integrated luminosity of about $2.5 fb^{-1}$ has been delivered from the Tevatron to the CDF II detector. The figures 2.5 and 2.6 show the development of the integrated and the peak luminosity since the start of Run II in 2002.

Parameter	Value for Run II
revolution frequency f [kHz]	47.713
number of protons per bunch N_p	$\sim 3 \times 10^{11}$
number of anti-protons per bunch $N_{\bar{p}}$	$\sim 9 \times 10^{10}$
number of bunches N_b	36
average beam size σ [μm]	~ 30
bunch length σ_l [m]	0.37
bunch spacing [ns]	396
beta function β^*	35
form factor F	0.7

Table 2.1: Accelerator parameters for the Tevatron Run II [43, 44]

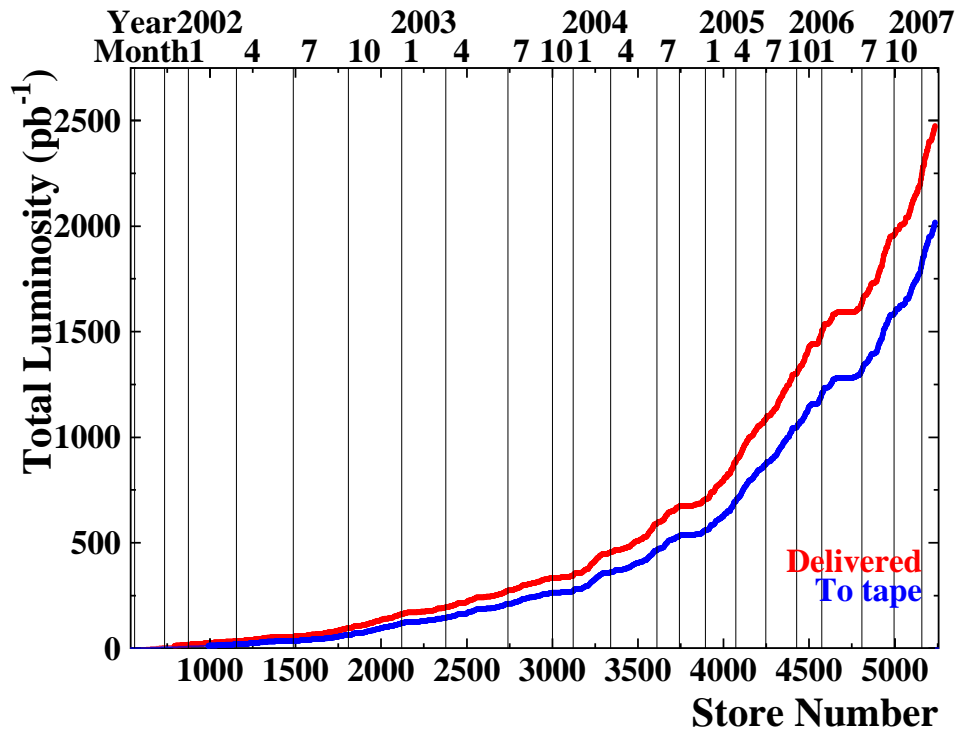


Figure 2.5: CDF II integrated Luminosity delivered from the Tevatron (red) and stored to tape (blue).

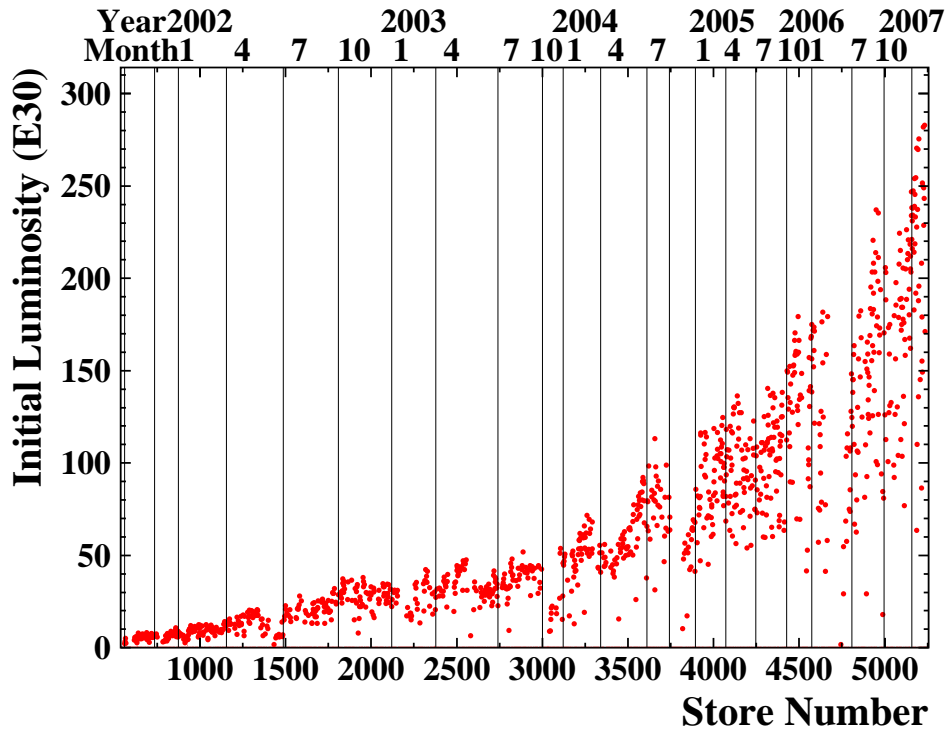


Figure 2.6: CDF II peak Luminosity.

2.2 The CDF Detector

The CDF II Detector (Collider Detector at Fermilab) is a multipurpose collider detector [45]. It detects particles produced in the $p\bar{p}$ -collisions and measures their properties. A cutaway view of the whole detector is displayed in figure 2.7. Within the solenoid the detector consists of a tracking system and a time-of-flight detector (TOF) for particle identification. A detailed view of the inner parts is shown in figure 2.8. Outside the solenoid, calorimeters and muon chambers are situated. The superconducting solenoid operates at 1.4 Tesla and the direction of its magnetic field is parallel to the beamline. The direction of the proton travelling along the beamline determines the z axis of the CDF coordinate system. Because the CDF experiment is left-right symmetric as well as cylindrically symmetric, one uses the azimuthal angle ϕ around the beamline and the polar angle θ from the beamline. Typically the pseudorapidity η is used instead of the polar angle θ :

$$\eta = -\ln \tan \frac{\theta}{2}$$

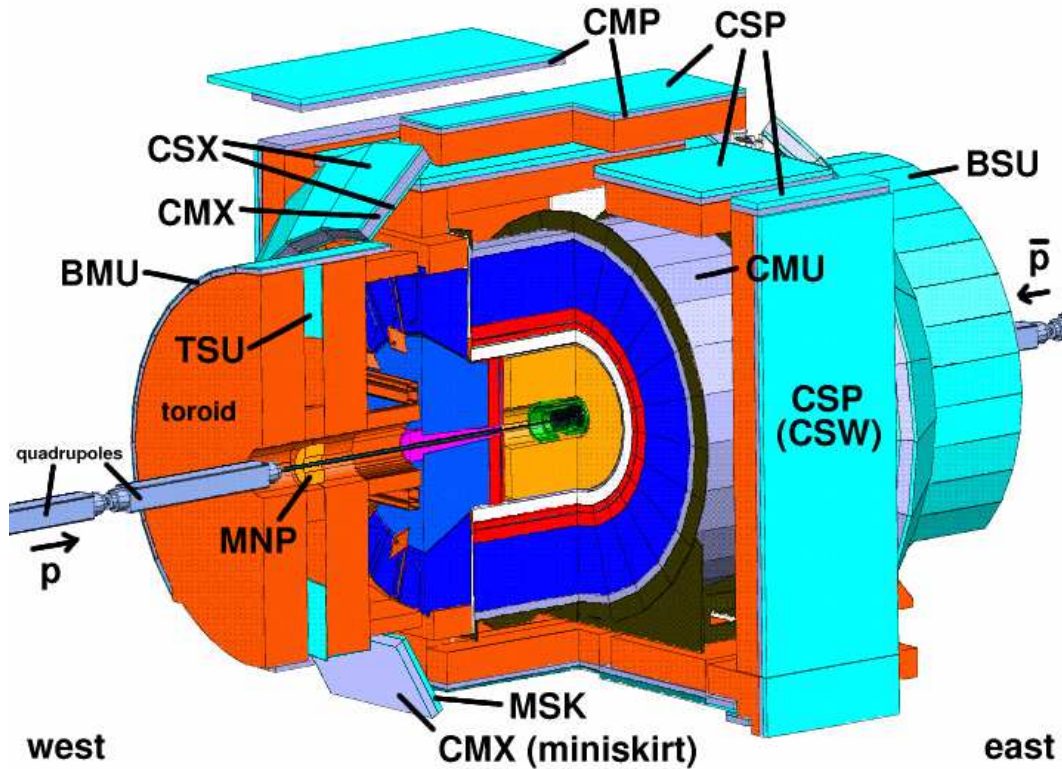


Figure 2.7: Cutaway view of the CDF II detector.

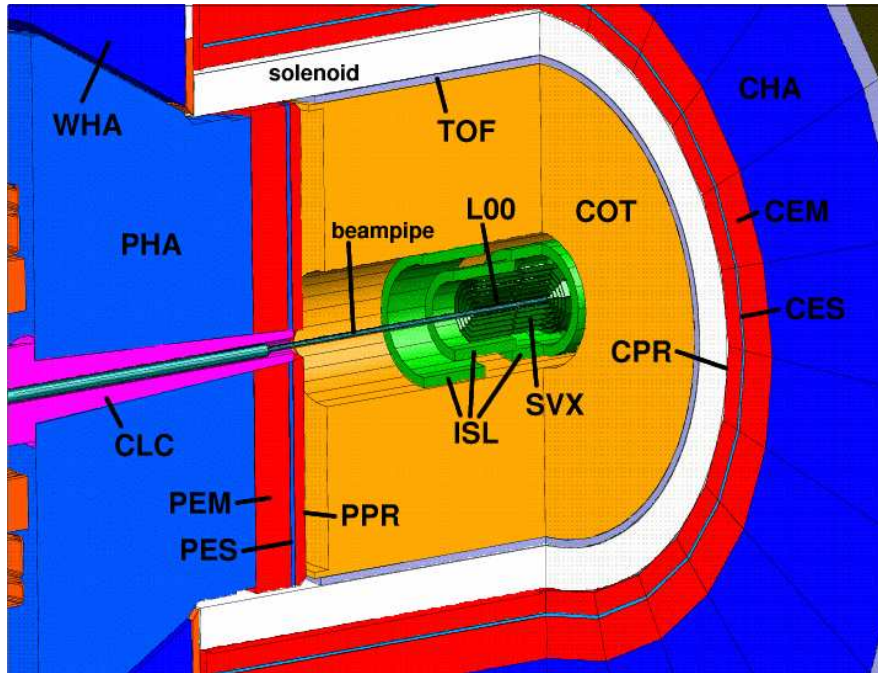


Figure 2.8: Cutaway view of the inner part of the CDF II detector. The acronyms are explained in the corresponding paragraphs of this chapter.

2.2.1 Tracking System

The tracking system of the CDF II detector is composed of four parts, the Layer 00 (L00), the Silicon Vertex Detector (SVXII), the Intermediate Silicon Layer (ISL) and the Central Outer Tracker (COT). To illustrate the η coverage of the tracking system a view of the r - z plane is shown in figure 2.9.

- Layer 00 : It is a single-sided silicon microstrip detector glued directly to the beampipe. The average radius is $r = 1.6$ cm, the length is about 80 cm and it has an η coverage of $|\eta| \leq 4.0$. Because of its closeness to the particle beam, it improves the B tagging efficiency and the precision of track measurements, especially the impact parameter resolution.
- SVX II : The SVX II serves both to measure tracks with a high precision and to detect secondary vertices at inner radii. It extends from a radius of $r = 2.1$ cm to $r = 17.3$ cm and is composed of three barrels with a length of 29 cm. These barrels cover a pseudorapidity range up to $|\eta| = 2$. The barrels each consist of five layers, which are divided into twelve wedges. All layers have axially aligned strips on one side to measure the ϕ coordinate. The layers 0, 1 and 3 have 90-degree stereo strips on the other side for a measurement of the z coordinate. The strips on layers 2 and 4 however use small-angle stereo strips instead of 90-degree stereo strips. Those strips are tilted only by 1.2° , which allows for a combination of the stereo and z coordinates to reduce the ambiguities

in the hit reconstruction. It ameliorates the resolution of the impact parameter and the azimuthal angle ϕ of tracks. Thereby the decay vertex of long lived unstable particles can be reconstructed with high spatial resolution.

- ISL : The ISL has a length of 195 cm. It consists of one layer located in the central region of the detector at $r = 22$ cm, and two layers in the forward/backward region at $r = 20$ cm and $r = 28$ cm. The central layer covers $|\eta| < 1$ and the forward/backward layers cover $1.0 < |\eta| < 2.0$. Both layers are double-sided silicon microstrip detectors designed the same way as the layers 2 and 4 of the SVX II. The main purpose of the ISL is to provide enhanced linking of tracks between SVX II and COT in the central region.
- COT [46, 47] : The outermost part of the tracking system is the Central Outer Tracker (COT), which is situated between $r = 40$ cm and $r = 137$ cm, and covers $|\eta| < 2$. It is a cylindrical drift chamber containing 2520 cells, divided into eight superlayers. The number of cells in each superlayer ranges from 168 for SL1 to 480 for SL8. Four superlayers have wires parallel to the beam (axial superlayers), whereas the other four have wires, that are tilted by 2 degrees with respect to the z-direction (stereo superlayers). Each cell has a wire plane made of 12 sense wires and 13 potential wires, made of gold plated tungsten, and two additional shaper wires at either end. Altogether there are 63000 sense and potential wires, extending over a length of 310 cm. The gas mixture Ar-Ethane-CF₄ (50:35:15) and the short drift distance of 0.88 cm enable the operation of the COT at drift times of about 100 ns. The shortness of the drift time is important to avoid the overlap of tracks from different beam crossings. A charged particle traversing the COT generates electron-ion pairs along its path. The electrons induce an electric signal in the sense wire, which is read out by an ASDQ (amplifier, shaper discriminator and charge encoding) chip [48]. Because of its large radial extension, the transverse momentum resolution is superior to the silicon detectors. In addition the density of tracks is smaller compared to the silicon detectors, leading to a better purity of the track reconstruction.

2.2.2 Particle Identification

The capability to distinguish between kaons, protons, electrons, muons and pions on a track level basis can be an essential input to many kinds of physics analyses. Two different types of measurements are placed at the disposal by the CDF II detectors within the solenoid. One measurement uses the energy loss of charged particles in matter, primarily caused by ionization or atomic excitation. The energy loss of particles in matter is described by the Bethe-Bloch equation, which is used for the parameterization of an universal curve in the COT, see figure 2.11. The universal curve then serves as an estimator of the energy loss for a given $\beta\gamma = \frac{p}{m}$. This estimate is compared to the measured energy loss and a probability for the assumed particle hypothesis is calculated. The energy loss of a particle in the COT is not measured directly. It is determined by measuring the pulse width of the signal in the COT. The pulse width is logarithmically dependent to the total charge deposited on the sense wire and therefore proportional to the energy loss :

$$\Delta t \sim \log Q \sim \frac{dE}{dx} \quad (2.1)$$

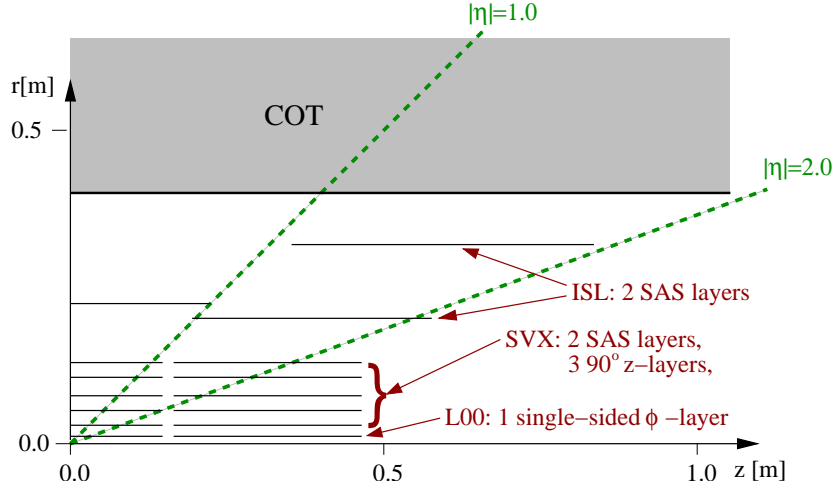


Figure 2.9: r-z view of a quarter of the tracking system

In general the energy loss of particles in the silicon detector could be used for particle identification. This has been studied in [49], but was never implemented due to the fact, that a good particle identification is only possible at very low momentum in the silicon detector. The other measurement utilizes the measurement of the particle's time of flight in the Time-of-Flight (TOF) detector. The TOF is located between the superconducting magnetic coil and the COT at $r = 140$ cm. It consists of 216 scintillator bars, which have a length of 280 cm. A photomultiplier is mounted at the end of each bar. The time-of-flight measurement of a particle requires information about the production time t_0 . Hence one has to match the measured TOF signal to a corresponding track and reconstruct the primary vertex position as well as the momentum of the track. Then the particle can be identified by calculating the invariant mass

$$m = \frac{p}{c} \sqrt{\frac{c^2(t_{TOF} - t_0)^2}{L^2} - 1} \quad (2.2)$$

with the pathlength L of the track calculated from the distance between the location of the primary vertex and the position of the measured TOF signal in the scintillator bar. In figure 2.10 the separation power of the TOF measurement for different particles and the separation power of COT dE/dx as a function of the track momentum is displayed.

2.2.3 Calorimeters

The CDF II calorimetry system is located outside the solenoid, see figure 2.8. It uses sampling calorimeters, which measure the particles and jet energies by absorbing all particles except muons and neutrinos. The system is divided into five independent calorimeters [46]:

- Central Calorimeters :
 - Central Electromagnetic Calorimeter CEM, Pb/scintillator sampling calorimeter, covers $|\eta| < 1.1$
 - Central Hadronic Calorimeter CHA, Fe/scintillator sampling calorimeter, covers $|\eta| < 0.9$

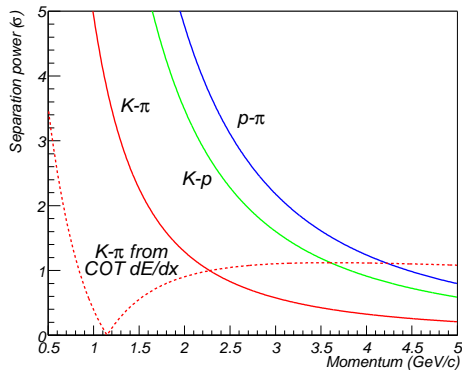


Figure 2.10: The TOF separation power for different particles (solid lines) compared to the COT dE/dx separation power (dashed line).

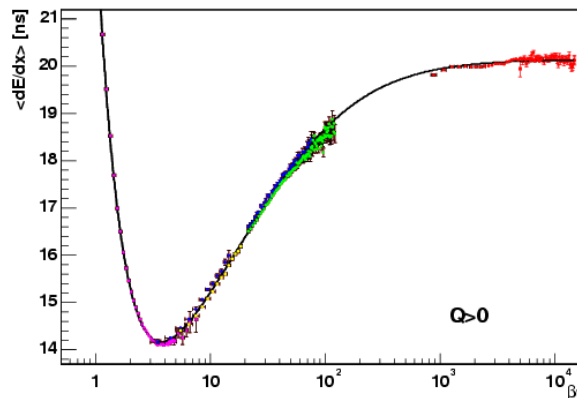


Figure 2.11: Universal curve of the COT dE/dx for positive tracks

- Endwall Hadronic Calorimeter WHA, Fe/scintillator sampling calorimeter, covers $0.8 < |\eta| < 1.2$
- End plug Calorimeters :
 - Plug Electromagnetic Calorimeter PEM, Pb/scintillator sampling calorimeter, covers $1.1 < |\eta| < 3.6$
 - Plug Hadronic Calorimeter PHA, Fe/scintillator sampling calorimeter, covers $1.2 < |\eta| < 3.6$

2.2.4 Muon Detector System

The muon system is the outermost part of the CDF II detector. It consists of four detectors made of scintillators and proportional chambers, that cover a region of $|\eta| < 2$. The track segment of muons measured in the muon chambers is called muon stub. This stub is matched to a track in the COT and then combined to form a muon candidate. The identification of muons is quite pure, because only a few kaons and pions reach the muon chambers. Muon fake rates are at a level of a few percent. An overview of the muon chamber properties is given in table 2.2.

	CMU	CMP	CMX	IMU
$ \eta $ -range	≤ 0.6	≤ 0.6	0.6-1.0	1.0-2.0
drift tube length [cm]	226	640	180	363
max. drift time [ns]	800	1400	1400	800
min. muon pt [GeV]	1.4	2.2	1.4	1.4-2.0

Table 2.2: Design parameters of the muon chambers [45, 46]

2.3 The Run II Trigger System

At the current state of operation of the Tevatron (36×36 bunches) the crossing rate is 2.5 MHz in the CDF II detector. This leads to an enormous data flow of ~ 0.5 TB/s, which is impossible to store since the maximum write speed of a tape is about 100 Hz and all available tapes ~ 2000 TB would be filled in about an hour. This necessitates a drastic data reduction by throwing out events which are not interesting for further physical studies. These events are called minimum bias events. The minimum bias events have a production cross section that is four (ten) magnitudes, higher than the production cross section for bottom quarks (top quarks). To increase the fraction of interesting events a preselection is essential. The preselection of interesting events is accomplished in the CDF II trigger system [50]. The trigger is assembled of three successive subsystems (Levels), which based on trigger tables decide if an event passes or not. The trigger tables contain trigger paths, which possess a set of rules defining requirements for the events. In the following section a short overview of the general properties of the CDF II trigger system is given. The data flow in the CDF II trigger system is shown in figure 2.12 and the mode of operation of the Level 1 and Level 2 trigger is displayed in figure 2.13. The Two Track Trigger (TTT), which is important for the exclusive reconstruction of B decays, is discussed in a later chapter of this thesis (see section 4.2.1.).

2.3.1 Level 1 Trigger

The Level 1 trigger uses information of all detector components, except the silicon vertex detector (SVX), read out after each beam crossing. The data from the SVX is kept in a pipeline for further processing in the Level 2 trigger. The other event information is used to form coarse versions of physics objects, referred to as trigger primitives. These primitives are identified by the eXtremely Fast Tracker (XFT), which reconstructs tracks by comparing the hit data to a list of pre-defined COT hit patterns. This gives a rough measurement of the transverse momentum p_T and the azimuthal angle ϕ of a track. In addition the XFT tracks are matched to muon chamber hits and calorimeter clusters. The time for a Level 1 trigger decision is about $5 \mu\text{s}$, which is longer than the time between two successive bunches. Therefore the information of the detector subsystems is stored in a pipeline at the front-end electronics. The limit of the acceptance rate of the Level 1 trigger is ~ 50 kHz.

2.3.2 Level 2 Trigger

The Level 2 trigger is an asynchronous system, which is able to store the information of up to four events in its buffer. Only about one event in 20000 passes the Level 1 and Level 2 trigger. The Silicon Vertex Tracker (SVT) is implemented in the Level 2 trigger and uses the information supplied by the Level 1 XFT as a seed. It extrapolates the XFT tracks to the SVX by adding silicon hit clusters and employing pattern-matching algorithms. The result are SVT tracks with measurements of the impact parameter d_0 , the transverse momentum p_T and the azimuthal angle ϕ . The resolution of the impact parameter d_0 is about $\sigma_{d_0} \sim 47 \mu\text{m}$ and is shown in figure 2.15. It is a combination of the transverse size of the beam profile and the intrinsic SVT impact parameter resolution. This good impact parameter resolution provides the possibility to distinguish between particles from the primary vertex and displaced vertices, as illustrated in figure 2.14. Thus one can enrich samples with long lived particles, e.g. b hadrons, by requiring impact parameters of $d_0 > 100 \mu\text{m}$ for a least two SVT tracks.

2.3.3 Level 3 Trigger

The Level 3 trigger is a software trigger installed on a special PC cluster, called L3 Farm. It receives data from the previous triggers with a rate of about 600-800 Hz. The Level 3 trigger performs a full reconstruction of the event and further improves the resolution of p_T , ϕ and d_0 . The L3 decision can be based on the full event topology, enabling cuts on L_{xy} , invariant masses etc.. The events accepted by the Level 3 trigger are first stored on disk in a storage area and later written to tape for analysis. At the end of the trigger chain about one event in 120000 survive all trigger requirements, leaving an output rate of ~ 100 Hz.

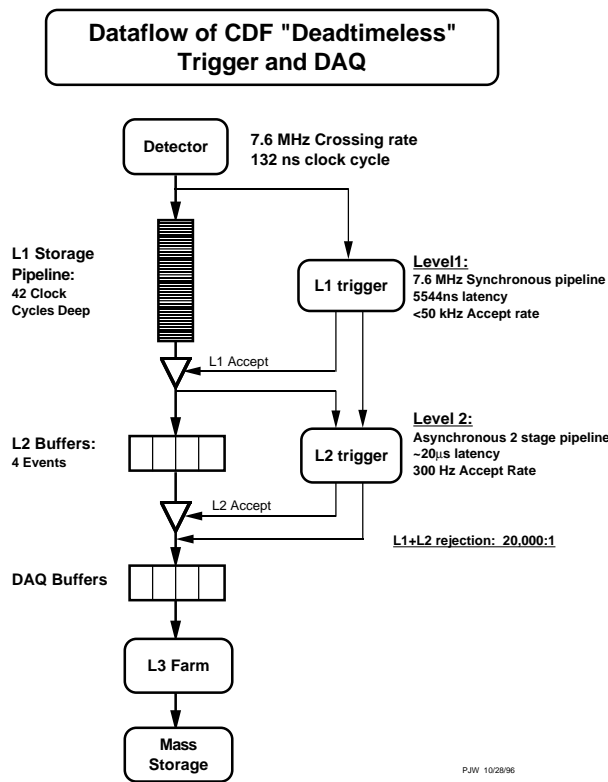


Figure 2.12: View of the dataflow of the CDF II Trigger System

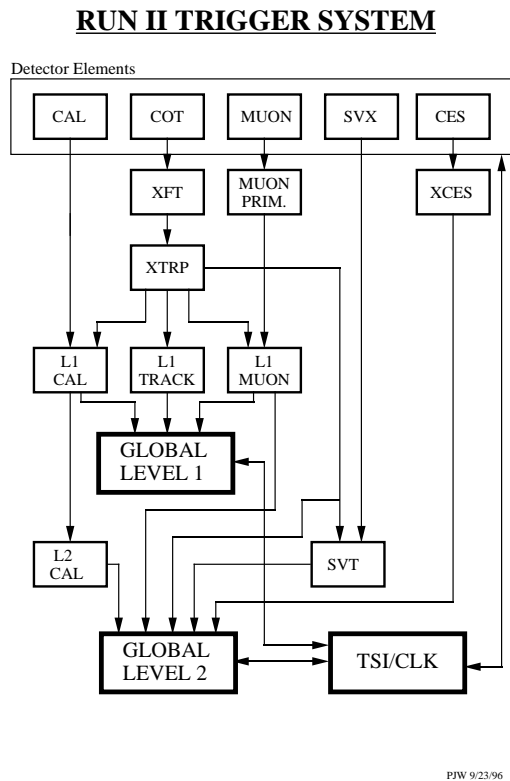


Figure 2.13: Diagram of the Level 1 and Level 2 Trigger of CDF II Trigger System

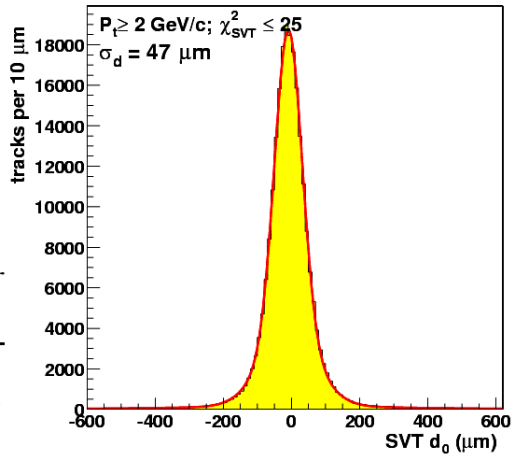


Figure 2.14: SVT impact parameter resolution including the $33\ \mu\text{m}$ beamspot

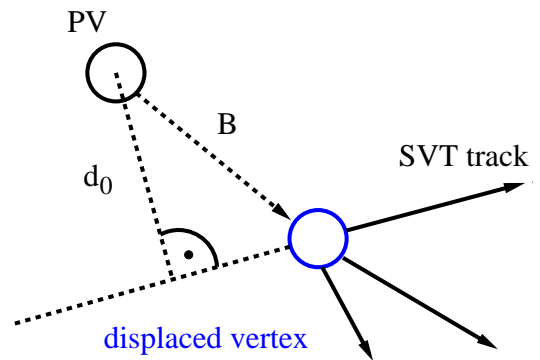


Figure 2.15: Definition of the track impact parameter d_0 for a SVT track

Chapter 3

Statistical Methods

3.1 Parameter Estimation

A common problem in particle physics and other quantitative sciences is the determination of parameters from measurements with errors. Ideally a measurement always gives exactly the same outcome if it is repeated under the same conditions. But this is not realized in nature, because we have to take effects like finite resolution of the experimental apparatus, thermal fluctuations or quantum effects into consideration, which cannot be controlled. Therefore one needs an optimal procedure to receive an accurate result for the parameter estimation from a given set of measurements. The criteria for these methods are consistency, unbiasedness, efficiency and robustness. One method commonly used for parameter estimation is the maximum likelihood method. This method is a very important component of this thesis and will be explained in detail in the following sections.

3.1.1 The Maximum Likelihood Method

Principles of the Maximum Likelihood Method [51]

Considering a set of n independent measurements $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ of the one- or multidimensional variable \vec{x} following a probability density $f(\vec{x}|\vec{a})$, with the unknown set of parameters \vec{a} , one can write the joint probability density function as

$$L(\vec{a}) = f(\vec{x}_1|\vec{a}) \cdot f(\vec{x}_2|\vec{a}) \cdots f(\vec{x}_n|\vec{a}) = \prod_{n=1}^n f(\vec{x}_i|\vec{a}) \quad (3.1)$$

The function $L(\vec{a})$ is called likelihood function and gives the probability to measure \vec{x}_i for a predefined set of parameters \vec{a} . The value $\hat{\vec{a}}$, that maximizes the function $L(\vec{a})$, is the best estimate for the parameter \vec{a} . This can be expressed by the condition

$$\frac{\partial L(a_k)}{\partial a_k} = 0 \quad \text{for all } k \quad (3.2)$$

It is advantageous to use the logarithm of $L(\vec{a})$, because the values of $L(\vec{a})$ can be very small and therefore cause numerical instabilities on computers. The logarithm is a monotonous function, thus leaving the minimum of $L(\vec{a})$ at the same position. Equation 3.2 can be rewritten as

$$\frac{\partial(\ln L(a_k))}{\partial a_k} = 0 \quad \text{for all } k \quad (3.3)$$

This leads to the definition of the negative log likelihood function

$$F(\vec{a}) = -\ln L(\vec{a}) = -\sum_{i=1}^n \ln f(\vec{x}_i|\vec{a}) \quad (3.4)$$

which has to be minimized to estimate \hat{a} . Sometimes the definition of $F(\vec{a})$ includes an additional factor of 2 to make it comparable to the method of least squares. During the process of minimization one has to take care that the probability density function $f(\vec{x}|\vec{a})$ is always normalized :

$$\int f(\vec{x}|\vec{a}) d\vec{x} = 1 \quad \text{for all } \vec{a} \quad (3.5)$$

The maximum likelihood method features several characteristics in the limit of an infinite number of measurements $n \rightarrow \infty$ [52, 53]:

- It is asymptotically consistent. The estimator \hat{a} of the parameter a converges to the true value a_0 :

$$\lim_{n \rightarrow \infty} \hat{a} = a_0$$

- It is asymptotically unbiased.
- It is asymptotically efficient. That implies that no unbiased estimator has a smaller variance than the estimator obtained with the maximum likelihood method.
- The function $L(\vec{a})$ approaches a multidimensional gauss function in the asymptotic limit and the variance of the estimator \hat{a} is zero.

Error calculation

In the case of the asymptotic limit one can expand the negative log likelihood function $F(\vec{a})$ around its minimum, where $\vec{a} = \vec{\hat{a}}$ and $\frac{dF}{da} = 0$, resulting in

$$F(\vec{a}) = F(\vec{\hat{a}}) + \frac{1}{2} \cdot \sum_{i,k} \frac{\partial^2 F}{\partial a_i \cdot \partial a_k} \cdot (a_i - \hat{a}_i)(a_k - \hat{a}_k) + \dots \quad (3.6)$$

$$= F(\vec{\hat{a}}) + \frac{1}{2} \cdot \sum_{i,k} G_{ij} \cdot (a_i - \hat{a}_i)(a_k - \hat{a}_k) + \dots \quad (3.7)$$

with $\vec{a} = (a_1, a_2, \dots, a_n)$ and \mathbf{G} the inverse of the covariance matrix \mathbf{V} of \vec{a} . The inverse covariance matrix \mathbf{V}^{-1} has the form of the Hesse matrix \mathbf{H} and can be used as an approximation of \mathbf{G} in the non-asymptotic case.

If there is only one parameter a , the evaluation of $F(a)$ around the minimum $a = \hat{a} \pm r \cdot \sigma$ in the asymptotic limit yields to

$$F(\hat{a} \pm r \cdot \sigma) = F(\hat{a}) + \frac{1}{2} r^2 \quad (3.8)$$

and the standard deviation σ arises from the comparison of $L(a)$ to the gauss function

$$\sigma(\hat{a}) = \left(\left. \frac{d^2 F}{da^2} \right|_{\hat{a}} \right)^{-\frac{1}{2}} \quad (3.9)$$

The case of many parameters can be discussed in a similar way. One defines the minimum of the function $F(\vec{a})$ with respect to all other parameters $a_j, j \neq i$

$$F_{min}(a_i) = \min F(\vec{a}) \quad (3.10)$$

The $r\sigma$ standard deviations then can be written as

$$F_{min}(\hat{a}_i \pm r \cdot \sigma) = F(\vec{\hat{a}}) + \frac{1}{2}r^2 \quad (3.11)$$

Deviations from the asymptotic behavior of the negative log likelihood function are approximated by asymmetric errors

$$F_{min}(\hat{a}_i + \sigma_r) = F(\vec{\hat{a}}) + \frac{1}{2} \quad (3.12)$$

$$F_{min}(\hat{a}_i - \sigma_l) = F(\vec{\hat{a}}) + \frac{1}{2} \quad (3.13)$$

The Maximum Likelihood Method for Binned Data

If there is a large number of measurements, it is useful to represent the frequency distribution of the data in a histogram. The x axis of a one-dimensional histogram is divided into J intervals, called bins. To simplify the following expressions a constant bin-width of $\Delta x = \frac{x_1 - x_0}{J}$ is assumed. The measured content of bin j is

$$n_j = \int_{x_0 + \Delta x(j-1)}^{x_0 + \Delta x j} \sum_{i=1}^n \delta(x - x_i) dx \quad (3.14)$$

with $\sum_{j=1}^J n_j = n$. Fitting the probability density $f(x|a)$ to the data points in each of the bins J leads to the fitted bin content

$$f_j(a) = \int_{x_0 + \Delta x(j-1)}^{x_0 + \Delta x j} f(x|a) dx \quad (3.15)$$

Assuming the entries n_j in each bin are random variables that follow a Poisson distribution with expectation value μ_j

$$P(n_j|\mu_j) = \frac{\mu_j^{n_j} e^{-\mu_j}}{n_j!} \quad (3.16)$$

$$\mu_j(a) = n \cdot f_j(a) \approx n \cdot f(x_0 + \Delta x(j - 0.5)|a) \cdot \Delta x \quad (3.17)$$

the likelihood, respectively the negative log likelihood function, can be written as

$$L(a) = \prod_{j=1}^J \frac{\mu_j^{n_j} e^{-\mu_j}}{n_j!} \quad (3.18)$$

$$F(a) = - \sum_{j=1}^J \ln \left(\frac{\mu_j^{n_j} e^{-\mu_j}}{n_j!} \right) = - \sum_{j=1}^J n_j \ln \mu_j + \sum_{j=1}^J \mu_j + \sum_{j=1}^J \ln n_j! \quad (3.19)$$

If the expected number of events in each bin is large, then we may approximate the Poisson distribution by a Gaussian distribution with variance $\sigma_j^2 = \mu_j$

$$P(n_j|\mu_j) = \frac{1}{\sqrt{2\pi\sigma_j}} e^{-\frac{(n_j-\mu_j)^2}{2\sigma_j^2}} \quad (3.20)$$

In this approximation the negative log likelihood takes the shape of a χ^2 distribution with k degrees of freedom. k is the difference between the number of bins J and the number of fitted parameters.

$$F(a) = -\sum_{j=1}^J \ln \left(\frac{1}{\sqrt{2\pi\sigma_j}} e^{-\frac{(n_j-\mu_j)^2}{2\sigma_j^2}} \right) = \frac{1}{2} \sum_{j=1}^J \frac{(n_j - \mu_j)^2}{\mu_j^2} + \text{const.} \quad (3.21)$$

To check the performance of a fit, one can use the ratio χ^2/k . This ratio should be $\simeq 1$ for a good description of the data points by the fitted function $f(x|a)$.

Minimization algorithms

The minimization of functions is a problem one has to face if one applies the maximum likelihood method to estimate fit parameters. This is a key issue in this thesis, because the maximum likelihood method is used in the determination of Δm_s as well as in the mass and lifetime fitting procedures. Throughout this thesis the fitting package MINUIT, implemented in ROOT, is used. The package MINUIT contains tools for the determination of the minimum value of a multi-parameter function and the analysis of the shape of this function around its minimum. The standard minimization algorithm is MIGRAD. It uses the variable metric method, which solves the minimization problem for a function $F(\vec{x})$, where the gradient vector $g(\vec{x}) = \nabla_x F(\vec{x})$ is available, but the Hesse matrix \mathbf{H} is not known explicitly. The algorithm updates the estimate of \mathbf{H} at each iteration using the changes on \vec{x} and \vec{g} made on that iteration. It is based on the following approximation

$$\vec{g}(\vec{x}_k + \Delta\vec{x}) \approx \vec{g}(\vec{x}_k) + \mathbf{H}_k \Delta\vec{x} \quad (3.22)$$

The direction of search $\Delta\vec{x}$ can be calculated from

$$\mathbf{H}_k \Delta\vec{x} = -\vec{g}_k \quad (3.23)$$

The next approximation of \vec{x} follows from a one dimensional minimization in the direction of $\Delta\vec{x}$

$$\vec{x}_{k+1} = \vec{x}_k + z_{min} \Delta\vec{x} \quad (3.24)$$

where z_{min} denotes the minimum of the one dimensional function $f(z) = F(\vec{x}_k + z \cdot \Delta\vec{x})$

Introducing the matrix \mathbf{U}_k one can rewrite the improved approximation of \mathbf{H}

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \mathbf{U}_k \quad (3.25)$$

The matrix \mathbf{U}_k can be chosen so that \mathbf{H}_{k+1} fulfils the relation

$$\mathbf{H}_{k+1}(\vec{x}_{k+1} - \vec{x}_k) = \vec{g}_{k+1} - \vec{g}_k \quad (3.26)$$

This can be expressed in terms of the vector $\vec{\delta}$ and $\vec{\gamma}$

$$\mathbf{H}_{k+1}\vec{\delta} = \vec{\gamma} \quad (3.27)$$

with $\vec{\delta} = \vec{x}_{k+1} - \vec{x}_k$ and $\vec{\gamma} = \vec{g}_{k+1} - \vec{g}_k$.

The relation quoted in equation 3.27 does not determine the matrix \mathbf{U}_k unambiguously. The most effective formula is the BFGS formula, named after Broydon, Fletcher, Goldfarb and Shanno. A detailed derivation of this formula can be found in [54–57]. The BFGS formula relates to :

$$\mathbf{H}_{k+1} = \mathbf{H}_k - \frac{(\mathbf{H}_k\vec{\delta})(\mathbf{H}_k\vec{\delta})^T}{\vec{\delta}^T\mathbf{H}_k\vec{\delta}} + \frac{\vec{\gamma}\vec{\gamma}^T}{\vec{\gamma}^T\vec{\delta}} \quad (3.28)$$

In the case of a one dimensional minimization, discussed in equation 3.24, the BFGS formula corresponds to

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\vec{g}_k\vec{g}_k^T}{\vec{g}_k^T\Delta\vec{x}} + \frac{\vec{\gamma}\vec{\gamma}^T}{z_{min} \cdot \vec{\gamma}^T\Delta\vec{x}} \quad (3.29)$$

The program package MINUIT [58, 59] contains not only tools for minimization but in addition tools to analyze the parameter errors. The error matrix calculated in MIGRAD yields usually good results. The error calculation in MIGRAD fails for non-linear problems or if the convergence to the minimum occurs too fast. An other tool to obtain the parameter errors is MINOS. MINOS is designed to work in non-linear cases and is a must if asymmetric errors are required, e.g. in case of non-parabolic behavior of the negative log likelihood function, see equations 3.12 and 3.13. Unlike MIGRAD, MINOS does not calculate the errors by using the curvature at the minimum and assuming a parabolic shape. But it needs a good minimum and a beforehand calculation of the error matrix as initial values. Therefore it is common to run MIGRAD prior to MINOS. The algorithm varies the parameter a_i with respect to all parameters \vec{a} in order to find the two values (positive and negative error) for which the minimum of the function $F(\vec{a})$ has the values $F(\hat{a}) + r$. The default value for r is $r = 1$, implying a 2σ asymmetric error. In this thesis usually the error matrix is obtained from MIGRAD, otherwise the use of MINOS is stated explicitly.

3.2 Artificial neural networks

Neural networks are sophisticated and powerful tools for problems in the domain of classification and prediction. The idea of artificial neural networks originates from researches in Artificial Intelligence, where it was attempted to model basic neural function of the brain. The brain consists of a very large number $\simeq 10^9$ of interconnected neurons. The neuron is a specialized cell, composed of an input structure (the dendrites), a cell body and an output structure (the axon), illustrated in figure 3.1. The dendrites of one neuron are linked by synapses to the axons of an other neuron. If the total signal received at the cell body from the dendrites exceeds a certain threshold, the neuron fires an electro chemical signal along the axon. The intensity of this signal depends on the synaptic efficiency. The layout of a biological neuron can be used as a blue print for an artificial neuron. The artificial neuron j , as displayed in figure 3.2, has a an input vector \vec{x} , corresponding to the dendrites in the biological neuron. The input is connected via weights w_{ij} , that have the same function as the synaptic efficiency.

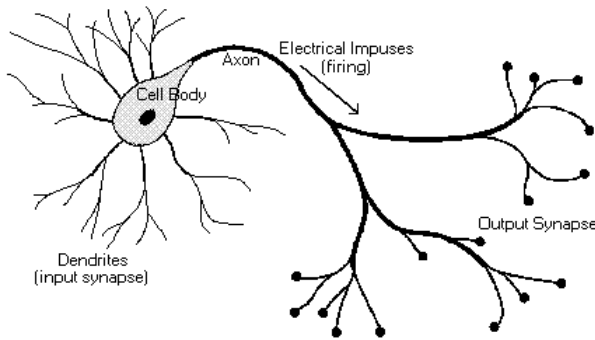


Figure 3.1: Sketch of the anatomy of a neuron

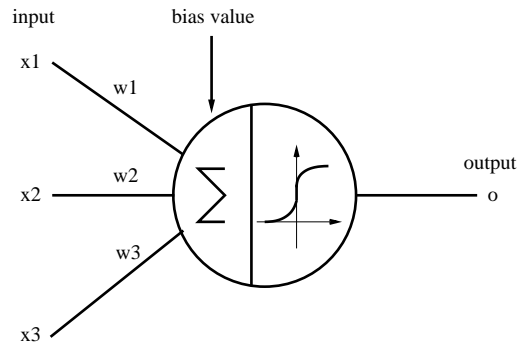


Figure 3.2: Layout of an artificial neuron

These weights are summed up in the neuron and an additional bias μ_j , controlling the threshold of the neuron, is subtracted. The output o_j of the neuron j is calculated by the activation function f . A commonly used activation function is the sigmoid function

$$S(x) = \frac{2}{1 + e^{-c \cdot x}} - 1 \quad (3.30)$$

It maps the interval $[-\infty, +\infty]$ to $[-1, 1]$ and has the property $-S(x) = S(-x)$. In figure 3.3 examples of the sigmoid for different values of c are shown. The model for an artificial neuron formulated in the last section can be expressed by the equation :

$$o_j = S\left(\sum_i w_{ij}x_i - \mu_j\right) \quad (3.31)$$

The single artificial neuron is already capable of performing the boolean AND, OR and NOT operations. It cannot be used as a XOR operator, because this operator is unlike the operators mentioned before not linearly separable. To overcome this problem one can arrange many artificial neurons, e.g. in a feed forward network.

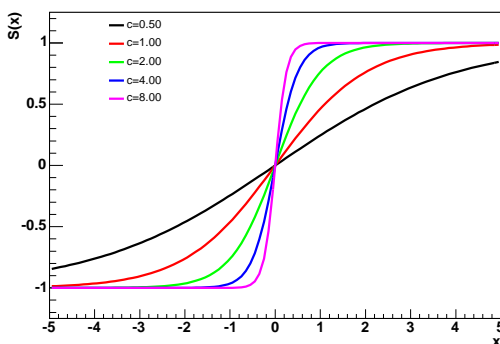
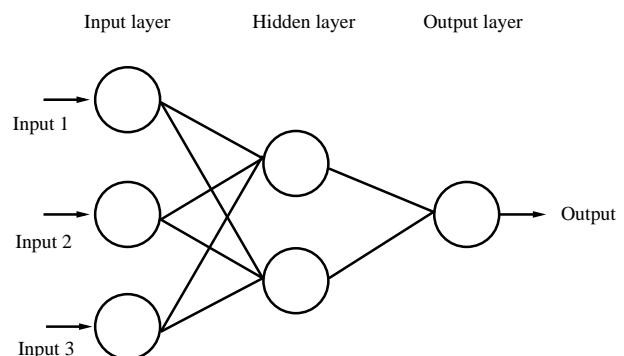
Figure 3.3: Examples of the sigmoid functions for different values of c 

Figure 3.4: Schematic view of a three layer feed forward network

3.2.1 The Feed Forward Neural Network

The feed forward neural network is one of many possible neural network topologies. It is called feed forward, because the information is only processed in the direction from input to output. The usual topology of the network consists of three types of layers : the input layer, the hidden layer and the output layer. Each layer consists of a distinct number of neurons, further referred to as nodes. The input layer has as many nodes as input variables are present. Each node of the input layer is linked to the nodes of the hidden layer. The choice of the number of nodes in the hidden layer is not completely arbitrary. If it has too many nodes, the network will be able to memorize everything by heart. Too few nodes will result in a network, that is not able to grasp all characteristics of a problem. In the case of a binary classification the output layer consists of only one node. The use of an additional bias node with a constant value of one in the input layer can enhance the performance of the network. A schematic view of the topology of a three layer neural network is illustrated in figure 3.4 . The network output o of a simple three layer feed forward network with the input vector \vec{x} , N input nodes and M nodes in the hidden layer, can be calculated via

$$o = S \left(\sum_j^M v_j S \left(\sum_i^N w_{ij} x_i - \mu_j \right) \right) \quad (3.32)$$

where v_j are the weights of the connection between the hidden layer to the output node, w_{ij} corresponds to the weights from the input layer to the hidden layer and μ_j is the bias value.

3.2.2 Training of a neural network

To train a network one needs training patterns, e.g. historical or simulated data, where the true classification is known. For a binary classification the true values can be e.g. one for signal or zero for background. The training relates to method of back propagation learning, in which the output is calculated for every input vector and compared to the true value. The comparison is done by calculating a cost function E , which describes the deviation of the desired value to the network output. The goal of back propagation is to minimize the cost function for all N training samples, by varying the weights connecting the nodes of different layers. Examples for cost functions are the squared difference of the true value and the network output

$$E = \sum_{i=1}^N (o_i - t_i)^2 \quad (3.33)$$

or the entropy function, which is used in NeuroBayes[®] [60]

$$E = \sum_{i=1}^N -\ln \left(\frac{1}{2} (1 + o_i \cdot t_i) \right) \quad (3.34)$$

where o_i and t_i are the network output, respectively the target value, of the i -th training pattern. The minimization of the cost function is a very complex task. This is due to the fact that the cost function depends strongly on the combination of weights and thus its surface has a complex shape in a multidimensional parameter space. In order to find the minimum

of the cost function, one can use the method of steepest descent [52]. Thereby the change of each weight is opposite to the direction of the cost function

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (3.35)$$

with η the learning rate. The learning rate determines how fast the weights are changed. If η is too big or if bad starting values for the weights are used, the algorithm will fail and not only the training of the neural network, but also the neural network itself will not be optimal.

3.2.3 Interpretation of the network output

The performance of a neural network can be evaluated by introducing the variables purity P and efficiency ϵ . The purity specifies the number of selected signal events compared to all selected events. Whereas the efficiency describes the ratio of selected signal event to all signal events in the whole sample. Then purity and efficiency can be expressed via the equations

$$P(\text{nnout}) = \frac{N_S(\text{>nnout})}{(N_S + N_B)(\text{>nnout})} \quad (3.36)$$

$$\epsilon(\text{nnout}) = \frac{N_S(\text{>nnout})}{N_S} \quad (3.37)$$

here N_S denotes the number of signal events and N_B the number of background events. nnout corresponds to the output of the neural network. To select a working point for the neural network, one usually plots the the purity versus the efficiency as displayed in figure 3.5. The optimal working point is the point closest to the coordinate (1, 1). If the network is trained optimally one can scale the network output to the interval [0, 1] and interpret it as probability. This assumption can be checked by plotting the purity versus the network output. It is correct, if the purity is a linear function of the network output, as shown in figure 3.6. One can prove this statement by writing down the mean contribution to the square of the cost function, see equation 3.33, for the input patterns with the neural network output nnout

$$E = P(\text{nnout}) \cdot (1 - \text{nnout})^2 + (1 - P(\text{nnout})) \cdot ((-1) - \text{nnout})^2 \quad (3.38)$$

The first term of this equation is the signal contribution with purity P and target value for the output of 1. The second term describes the background contribution with purity $(1-P)$ and a target of -1. The training is optimal if the cost function is minimal, leading to

$$\frac{dE}{d(\text{nnout})} = 0 \rightarrow P(\text{nnout}) = \frac{(\text{nnout} + 1)}{2} \quad (3.39)$$

3.2.4 NeuroBayes[®]

NeuroBayes[®] [60] is a sophisticated software for the work with neural networks. It is based on Bayes' theorem and consists of a three layer feed forward neural network. It was developed at the University of Karlsruhe and further evolved at the spin off company $\langle\text{phi-t}\rangle^{\text{®}}$. The program package provides tools for binary classification problems as well as for the prediction of probability density functions. All neural networks specified in this thesis are based on the NeuroBayes[®] package. In this section the basic features of NeuroBayes[®], that are important to a binary classification, are described.

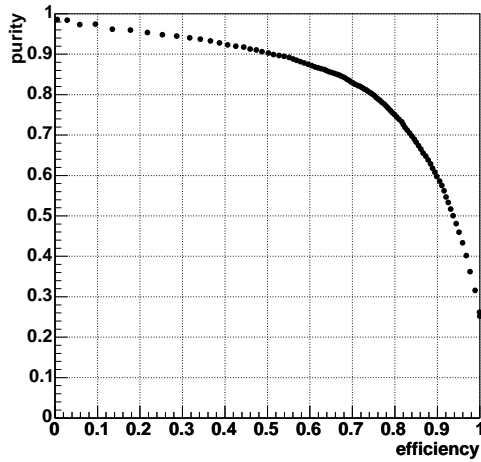


Figure 3.5: Example of a purity-efficiency plot to determine a working point.

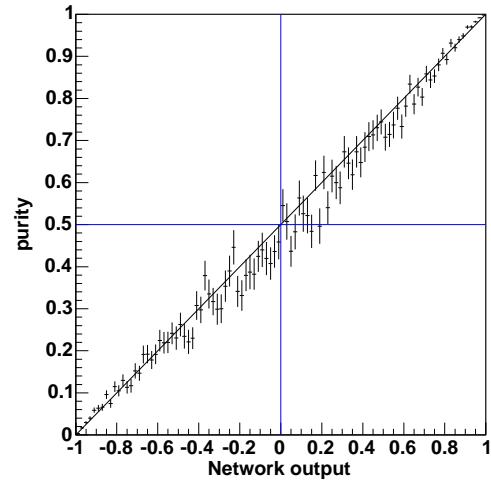


Figure 3.6: The signal purity plotted against the output of the neural network.

Bayes' theorem

Bayes' theorem [61] states the following relation, connecting the conditional probability of B, given A, $P(B|A)$, with the conditional probability of A, given B, $P(A|B)$.

$$P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)} \quad (3.40)$$

where $P(A)$ is the prior probability of A, which is independent of the prior probability $P(B)$ of B. This can be expressed as well in a physical context. The a posteriori probability $P(A|B)$ then denotes the probability of a theory for the observed data, and $P(A|B)$ is the likelihood to measure the data assuming the theory. In the case of the binary decision of a neural network, its output can be interpreted as a Bayesian a posteriori probability. In addition NeuroBayes[®] is able to make an event-by-event prediction of the full a posteriori probability density function.

Preprocessing of Variables

The preprocessing of input variables is a method, which can help in locating the minimum of a cost function. In the NeuroBayes[®] package a global and an individual preprocessing is implemented. The global preprocessing at first flattens the distribution of the input variables, maps them to the interval $[-1,1]$ and then transforms them into Gaussians with mean 0 and width 1. Because of correlations of the variables to the target as well as to themselves, the variables are decorrelated by rotating the covariance matrix until it is a diagonal unit matrix. The individual preprocessing treats special properties of single variables. Examples for the use of individual preprocessing are

- input distributions with a δ function or in case of events with undefined variables
- discrete variables that can be divided into ordered or unordered classes

- the replacement of input distributions with statistical fluctuations and non linear correlations to the target by a spline fit

Furthermore the significances of the input variables are calculated during the preprocessing. This calculation is done iteratively. At first one variable after the other is removed and the loss of total correlation is ascertained. In the next step the variable with the smallest loss of information when omitted is eliminated. This is repeated until all correlations to the target are determined. Finally one obtains a list of significance of the input variables to the target. A detailed description of this method can be found in [5].

Regularization

There are several methods of regularization to optimize the training of a neural network. One method is to reduce the number of free parameters by removing insignificant connections or nodes during the training and therefore enhancing the signal to noise ratio. Another method is to add a penalty term to the cost function in order to suppress oscillations around minima. The new cost function can be written as

$$\tilde{E} = E + \tau P = E + \tau \cdot \frac{1}{2} \sum_i w_i^2 \quad (3.41)$$

The penalty term punishes the neural network for large weights resulting in a network with smaller weights. This is also called weight decay, because in absence of any input the weights will tend to decrease exponentially. The benefits of this method are the reduction of the impact of random fluctuation of the input data and the prevention of oscillatory behavior of the network.

Chapter 4

Analysis Outline

4.1 Overview of the Analysis

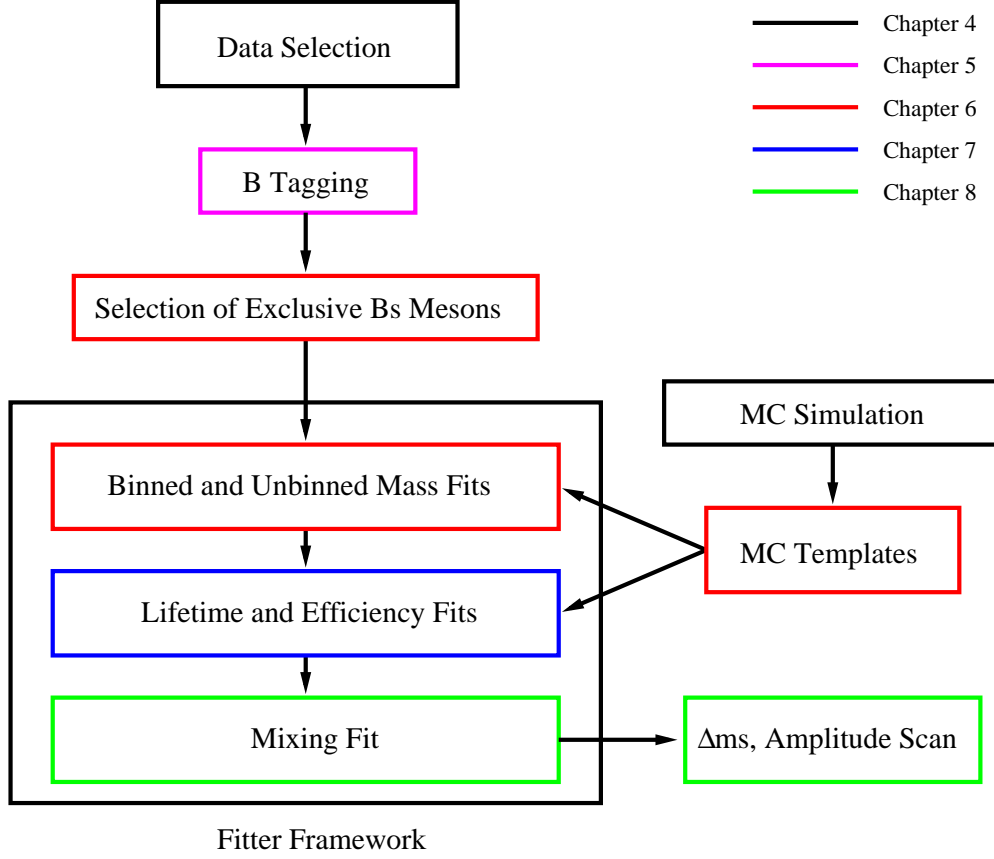
The time-dependent measurement of the B_s oscillation is, due to the large mixing frequency of the B_s meson, a challenging task. This chapter gives an overview of the ingredients necessary for the determination of Δm_s . There are several prerequisites one has to consider :

- One has to identify data containing long lived B mesons. This is achieved by using the Two Track Trigger for the preselection of data.
- The flavor of the B_s meson at production time needs to be determined. Therefore various tagging algorithms are applied to the data.
- The B_s meson decaying into a flavor eigenstate has to be reconstructed in order to obtain the flavor of the B_s at the decay time.
- The proper decay time of the B_s meson has to be measured.

The ingredients for the Δm_s measurement listed above form the basic structure of the following chapters of this thesis as summarized in figure 4.1. To finally determine the mixing frequency, one combines all available information in a likelihood function. The likelihood function for an event has the following fundamental form :

$$L = (1 - f_b) \cdot P_S(\xi, t, \sigma_t) \cdot P_S(m) + f_b \cdot P_B(\xi, t) \cdot P_B(m) \quad (4.1)$$

with $f_b = \frac{N_B}{N_B + N_S}$ denoting the fraction of background events. The subscript S, respectively B, mark the signal and the background contributions to the likelihood function. ξ denotes the tagging decision. t is the proper decay time and σ_t the decay time resolution. m is the reconstructed mass of the B_s meson candidate. The terms $P_S(\xi, t, \sigma_t)$ and $P_B(\xi, t)$ describe the probability density functions of the lifetime space including the mixing term and $P_{S,B}(m)$ corresponds to the description of the mass space. It is crucial for this analysis to model the constituents of this likelihood function as accurately as possible. The composition of the likelihood is the focus of the chapters 6 to 8. The cornerstone of the mixing analysis is the fitting framework which provides the methods for the binned and unbinned fits of various quantities, as well as for the mixing fit.

Figure 4.1: Road-map of the Δm_s measurement.

4.1.1 Mixing probability

The probability of a B_s meson oscillating into a \bar{B}_s meson at a given time t is the initial point of this analysis. It can be derived from equation 1.4 to be

$$P_{B_s \rightarrow \bar{B}_s} = \frac{\Gamma}{2} e^{-\Gamma t} (1 - \cos(\Delta m_s t)) \quad (4.2)$$

with the assumptions, that $|\eta_m|^2$ is one and the lifetime difference of the B_H and B_L is zero :

$$\Delta\Gamma = 0 \rightarrow y = 0 \rightarrow \cosh yT = 1$$

The time dependent asymmetry (equation 1.5) then simplifies to

$$a(T) = \cos \Delta m_s t \quad (4.3)$$

The mixing probability has to be modified to take into account the effects of the experimental resolution of the proper decay time measurement and performance limitations of the flavor tagging algorithms. These modifications are introduced in chapter 8 of this thesis at the point of the detailed description of the mixing term of the likelihood function.

4.1.2 Significance of the Δm_s Measurement

The significance \mathcal{S} of the Δm_s measurement can be calculated from the Fourier transformation or the maximum likelihood method. It can be expressed via the following formula

$$\mathcal{S} = \sqrt{\frac{\epsilon D^2}{2}} e^{-\frac{(\Delta m_s \cdot \sigma_{ct})^2}{2}} \frac{N_S}{\sqrt{N_S + N_B}} \quad (4.4)$$

where D denotes the dilution of the tagger and ϵ the efficiency of the tagger defined in section 5.1.1. σ_{ct} is the decay time resolution. N_S is the number of signal events and N_B the number of background events.

The expression above can be divided into three parts, that mirror the contributions of the tagging, the lifetime resolution to the Δm_s measurement and the selection of exclusively reconstructed B_s mesons. The first part is proportional to the square root of the tagging power $T = \epsilon D^2$. This quantity is the figure of merit for a tagger. It denotes the statistical effectiveness of a tagger. The second part of the significance takes the event-by-event resolution of the proper decay time into account. The contribution of this term increases with the mixing frequency. The last term describes the significance of the selected exclusively reconstructed B_s meson. The methods to enhance this term are subject to chapter 6 of this thesis.

4.2 Data Selection and Event Reconstruction

The data used for this thesis was recorded between February 2004 and February 2006 at the CDF II Experiment. The total integrated luminosity nearly reaches 1 fb^{-1} . In table 4.1 the three datasets forming the complete data are listed. All data were required to pass the Two Track Trigger and the good run selection. The good run selection is a list containing flags set by the shift crew during the data taking and after the completion of a run. These flags show whether all detector components and triggers were working according to their parameters during this run. This analysis makes use of the good run list provided by the DQM group [62]. For this analysis the requirements of this list are basically a working tracking system and SVT trigger system.

dataset (CDF internal notation)	data acquisition period	integrated luminosity [pb^{-1}]
xbhd0d	Feb 2002 to Aug 2004	341
xbhd0h	Dec 2004 to Sep 2005	397
xbhd0i	Oct 2005 to Feb 2006	253
all datasets		991

Table 4.1: Summary of the datasets used in this analysis.

4.2.1 The Two Track Trigger

The Two Track Trigger is the core part for the selection of the data used for this analysis. It relies on the track information provided by the XFT and the SVT. The basic principle of the Two Track Trigger is the selection of a track pair displaced from the primary interaction vertex. The requirements for an event to pass the Two Track Trigger are :

- a minimum of two XFT tracks with transverse momentum $p_t > 2$ GeV and the difference in the angle ϕ of both tracks $\Delta\phi^{1,2} < 135^\circ$ at the Level 1 trigger stage.
- a least two SVT tracks with $p_t > 2$ GeV, $100 \mu\text{m} < |d_0| < 1$ mm, $\chi_{SVT}^2 < 25$, $L_{xy} > 200 \mu\text{m}$ and $2^\circ < \phi^{1,2} < 90^\circ$ are required at Level 2 of the trigger chain.
- at Level 3 the Level 2 trigger decision are confirmed by the offline reconstruction.

In order to meet the demands of taking data in different luminosity scenarios, additional requirements are introduced at the Level 2 trigger stage. This is implemented by the Two Track Trigger subpaths :

- B.LOWPT : the low- p_t scenario
at least two SVT tracks with $p_t^1 + p_t^2 > 4.0$ GeV
- B.HIGHPT : the high- p_t scenario
at least two oppositely charged SVT tracks with $p_t^1 + p_t^2 > 6.5$ GeV
- B.CHARM : at least two SVT tracks with $p_t^1 + p_t^2 > 5.5$ GeV

In periods of high luminosity it may be necessary to prescale these trigger paths. This decreases the deadtime of the trigger system and frees bandwidth for other trigger paths. The prescaling of trigger paths consists in accepting one out of $N = 1/p$ events that have passed the trigger requirements. The factor p is the prescale factor, which is a dynamical quantity adjusted according to the luminosity. It is stored in a data base further accessible via the offline software.

4.2.2 Event Reconstruction

The data delivered by the Two Track Trigger can not yet be used for a high level physics analysis. It has to be processed by reconstruction algorithms provided by the CDF II offline software.

Tracking

At first tracks have to be reconstructed from the measurement of hits in the tracking system of the CDF II detector. There are several different tracking algorithms implemented in the CDF II offline software. The following list of algorithms corresponds to the sequence of their usage in the CDF II offline software :

- COT tracking : The algorithm reconstructs tracks of particles that have traversed the entire drift chamber. The reconstructed tracks have a high purity, because the track density in the drift chamber is low. Therefore only a small number of duplicate tracks are reconstructed. The COT tracking algorithm is limited to tracks of a minimum transverse momentum of 0.5 GeV and an $|\eta| < 1$.

- Outside In (OI) tracking : The OI algorithm extrapolates the tracks from the COT back to the silicon and adds corresponding hits along the reference trajectory. Because COT tracks are used as seed, this algorithm has the same geometrical limitations as the COT tracking. The advantage of the OI algorithm is the gain in the impact parameter and z position resolution compared to the pure COT tracks.
- Silicon Standalone (SISA) tracking : The SISA algorithm uses exclusively silicon measurements to reconstruct tracks. It is able to reconstruct lower transverse momentum tracks $p_t > 0.2$ GeV than the OI and COT algorithms. Furthermore tracks within the full coverage of the silicon detector $|\eta| \leq 2$ are reconstructed. The drawback of this algorithm is the requirement of two silicon 3D hits in the four available layers. This decreases the efficiency of the algorithm due to holes in the coverage of the SAS layers.
- Silicon Forward (FWD) tracking : The FWD algorithm uses a similar strategy as the SISA algorithm. It reconstructs tracks from one 3D hit and one r - ϕ hit instead requiring two three dimensional silicon hits. To reduce the combinatorics and the CPU time usage in the track reconstruction, the minimum transverse momentum is set to be larger than 0.8 GeV.
- Inside Out (IO) tracking : The IO algorithm extrapolates tracks from the silicon detector into the COT. In comparison to the SISA and FWD algorithms the IO algorithms improves the momentum resolution and decreases the fake track rate. This is due to the increased track length, the additional measurements in the COT and the confirmation of the silicon tracks by the COT.

A detailed description of the tracking algorithms can be found in [63].

The raw tracks reconstructed by the tracking algorithms are then used to calibrate the magnetic field and the material description of the detector on the invariant mass of the J/ψ . First the GEANT material map used for the track fitting [64] is corrected. Then the value for the magnetic field used in the calculation of the track momentum from the track curvature is calibrated on the world average value for the J/ψ mass. This procedure removes the momentum dependence of the reconstructed J/ψ mass [65]. Furthermore the COT covariance matrix is rescaled because of the underestimation of the measurement errors of the track reconstruction in the COT. After the rescaling the tracks are refitted according the particle hypothesis expected in the reconstructed final state. The refit of the tracks picks up the rescaled COT tracks and adds the according hit measurements of L00 of the silicon vertex detector. A Kalman fitter implemented in the KalTrack package [66] is usually used for the track refit.

Reconstruction of a B_s event

The package BottomMods [67, 68] of the CDF II offline software is used to reconstruct a B_s meson. The package provides a modular structure which can be easily adapted to reconstruct different decays. The input of the first module usually consists of low level objects like reconstructed tracks. The output of each module is propagated to the input of the next module in the reconstruction chain. The output of the last module of the reconstruction chain is a list of B_s candidates for an event. The first step of the reconstruction requires tracks that posses a successful helix fit and a physical error matrix for the track parameters. The next step is the refitting of tracks with either a pion or kaon mass hypothesis.

These tracks are stored as collections of stable particle candidates and serve as an input for the reconstruction of unstable particles. The vertex fit for all necessary vertices in the examined decay chain uses the CTVMFT package [69]. It calculates the vertex positions, the invariant masses, the four momenta of all involved particles and the corresponding covariance matrices. To reduce the computation time of the time consuming vertex fit, a soft preselection on the estimated invariant mass of a track pair can be applied before the vertex fit. After performing the vertex fit one can add further constraints to reduce the number of the selected B_s candidates. At this stage the trigger requirements have to be fulfilled as well. The information of the reconstructed B_s candidates is stored in the BStntuple framework. This framework contains all information of the vertex fits obtained during the reconstruction of the B_s candidates for a given event. In order to reduce the size of the Two Track Trigger dataset, it is converted to the ROOT Ntuple format and only the B_s decays one is interested in are selected. Furthermore the use of ROOT Ntuples is independent of the CDF II software and allows to analyze the data in the widely used ROOT framework.

4.3 Monte Carlo Simulation

The analysis of B_s oscillations requires simulated B events as an input in many situations. The training of the neural networks used in the selection of exclusively reconstructed B_s mesons depends on a reliable description of B_s meson decays. Furthermore the templates of the background shapes for the binned and unbinned mass fits are derived from simulated events. Finally one obtains the efficiency function for fits in the lifetime space from simulated signal events.

To generate the necessary Monte Carlo samples one can chose between two event generators [70]. The PYTHIA [71] event generator creates $b\bar{b}$ pairs via several production mechanisms. The most prominent are flavor creation, flavor excitation and gluon splitting processes [72]. The corresponding Feynman graphs for these processes are displayed in the figures 4.2, 4.3 and 4.4. To optimize the speed of the generation of simulated events, the underlying event structure of each event is fixed while the b-hadrons are redecayed into selected final states. To redecay the b-hadrons the EvtGen program package [73] is used. In order to simulate the detector one uses the cdfSim software [74]. After the detector simulation the simulated data has the same structure as real data. The simulation of the Two Track Trigger is handled by the program TrigSim++ [75], which includes amongst others the simulation of the SVT. The simulated events then undergo the same reconstruction procedure as real data described in section 4.2.2.

Another possibility of obtaining simulated events is to use BGEN as an event generator [76]. In contrast to PYTHIA the events simulated by the BGEN event generator only consists of a b-hadron without the opposite b quark and the fragmentation tracks. The advantage of BGEN is the possibility to obtain large Monte Carlo samples in a short period of time. The BGEN Monte Carlo samples are mainly used for analyses where no opposite b quark or fragmentation tracks are needed.

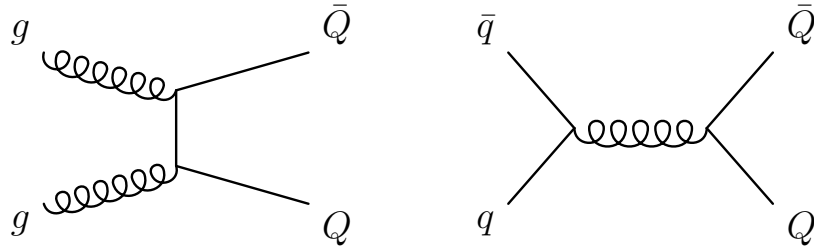


Figure 4.2: Leading order diagrams for heavy-flavor production.

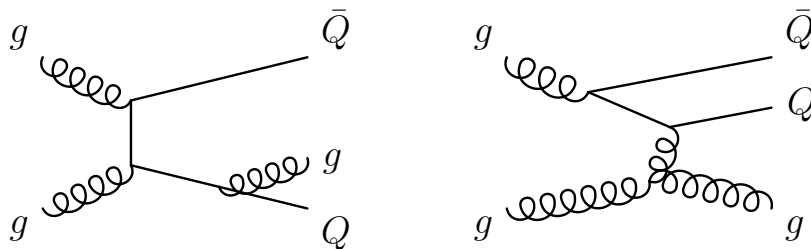


Figure 4.3: Pair creation with gluon emission (left) and flavor excitation (right) diagrams .

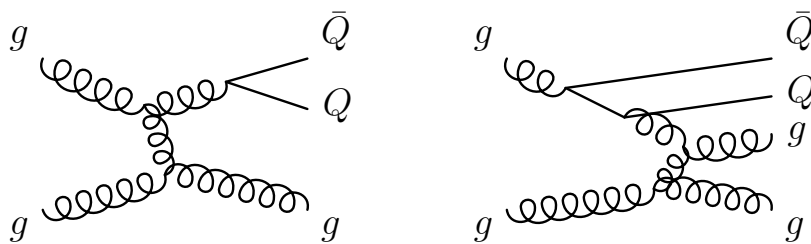


Figure 4.4: Gluon splitting diagram (left) and events classified as gluon splitting but of flavor-excitation type (right).

4.4 The Fitter Framework

The underlying concept of the fitter framework is to provide a flexible structure applicable for many analysis. It is designed in a modular concept, so one does not need to edit the entire fitter code for different purposes but only has to include additional modules. An overview of the basic composition of the fitter framework is given in figure 4.5.

In case of the B_s oscillation analysis, the fitter framework mainly serves as a general tool to facilitate the handling and bookkeeping of the fit functions used for the various decay modes. Because of the six decay channels studied in this analysis, it is necessary to keep track of nearly 500 parameters for the combined mixing fit. Therefore one has to take care of the transfers of fitting parameters between the Mode Class, the Fitter and MINUIT. In the ModeClass the methods for the likelihood functions as well as for the binned fits of the mass spectra are contained. Because the ModeClass includes all properties of the decay mode, the creation of a supplementary ModeClass is sufficient for the addition of other decay modes to the fitter framework. In order to simplify the fitting procedure a single Mode Class valid for all studied decay channels is used in this analysis. The AbsMode Class is the common interface for all Mode Classes to the Fitter.

The Fitter itself reads in the parameters for each mode and copies them to a map suited for the process of minimization in MINUIT. The name of each parameter serves as a key for the map and is the identifier for the parameter in MINUIT. If parameters have the same name in different modes, they are regarded as a common parameter for the minimization. This provides a flexible structure in which one can decide whether a parameter should be common or not. Examples for common parameters are the mixing frequency or the lifetime of the B_s meson. MINUIT then minimizes the likelihood requested by the steering executable and returns the fit result to the Fitter. The Fitter copies the parameters obtained from MINUIT back to the original parameter list and thereby restores the previous order of the parameters. This process is illustrated in figure 4.6. This framework is currently used not only for the B_s mixing analysis but as well for the analysis of the B_s^{**} , the B^{**} and further for the $\Delta\Gamma$ analysis in the decay $B_s \rightarrow J/\psi\phi$.

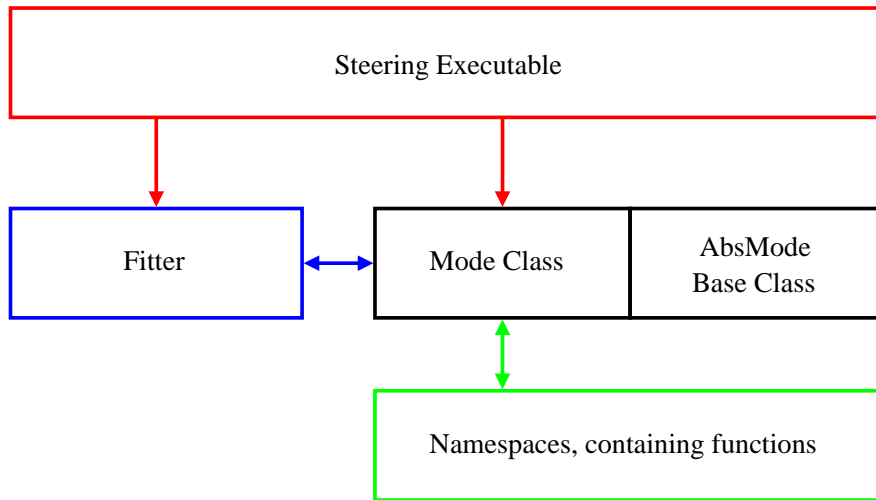


Figure 4.5: Overview of the components of the fitter framework.

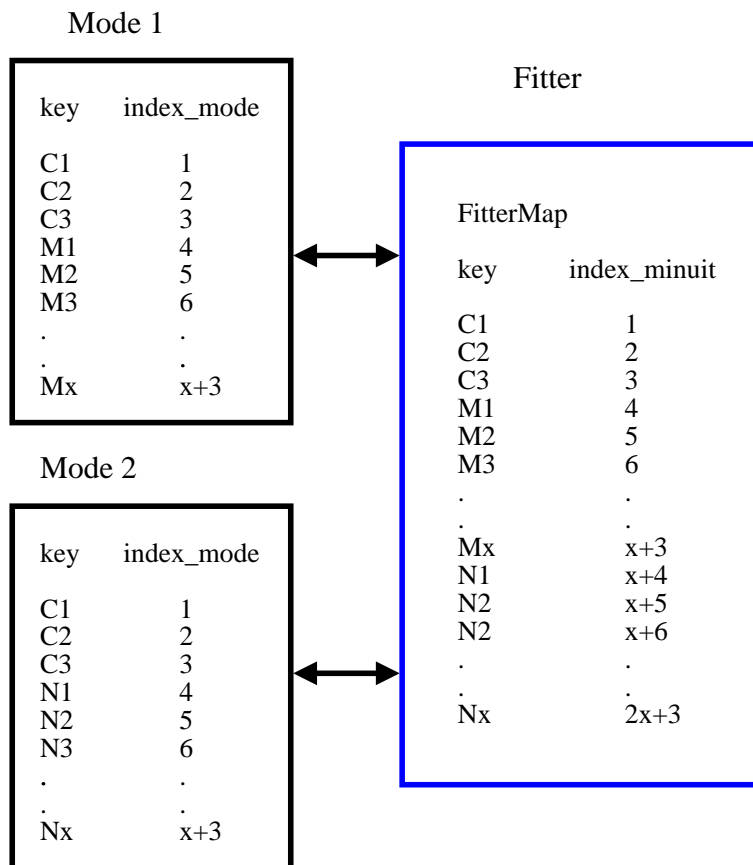


Figure 4.6: Illustration of the transfer of parameters between the Mode and the Fitter. The label C denotes common parameters, the label M and N the parameters of the different modes.

Chapter 5

B Flavor Tagging

5.1 General Properties of B Flavor Tagging

A crucial ingredient not only for the measurement of Δm_s , but as well for the measurement of mixing-induced CP violation in the B meson system is the B flavor tagging. It is used to determine the flavor of the B meson at production time. In figure 5.1 a schematic view of a B_s event is displayed. This event is divided into two hemispheres, the same side and the opposite side. The same side is defined by the reconstructed B_s meson. The opposite site contains the remaining b quark, which can hadronize into a B_s^0 , a B^0 , a B^+ or b baryons. The performance of the flavor tagger is quantified by the dilution and the efficiency.

5.1.1 Dilution and Efficiency

The dilution describes how often the tagging decision returned by the tagging algorithm is correct. It is 0 for a random tagging decision and 1 for a perfectly working tagger. One can define the dilution as the difference between the number of correct tagging decisions N_{RS} and wrong decisions N_{WS} normalized to all tagging decisions.

$$D = \frac{N_{RS} - N_{WS}}{N_{RS} + N_{WS}} = \frac{2 \cdot N_{RS}}{N_{RS} + N_{WS}} - 1 = 2 \cdot P_{RS} - 1 \quad (5.1)$$

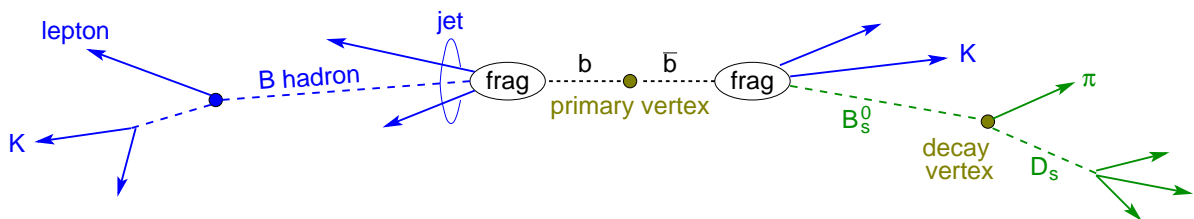


Figure 5.1: Schematic view of a B_s event. The same side hemisphere is defined by the fully reconstructed B_s meson (right part of the plot). The opposite side hemisphere contains the hadrons and jets coming from the fragmentation process of the remaining b quark (left part of the plot).

where P_{RS} is the probability for a correct tagging decisions.

The efficiency of a tagger can be expressed by the ratio of events with a tagging decision to all events.

$$\epsilon = \frac{N_{RS} + N_{WS}}{N_{RS} + N_{WS} + N_{NT}} \quad (5.2)$$

with N_{NT} the number of events, where no tagging decision is given by the tagging algorithm.

5.1.2 Opposite Side Tagging

The b flavor of the opposite side B meson can be identified by several different methods, like kaon, lepton, jet charge or vertex charge tagging. The main problem of opposite tagging arises from the fact that the hadronized B meson if neutral can oscillate into its anti particle. This affects taggers that determine the flavor of the b at decay time. Further problems are the separation into hemispheres and the acceptance.

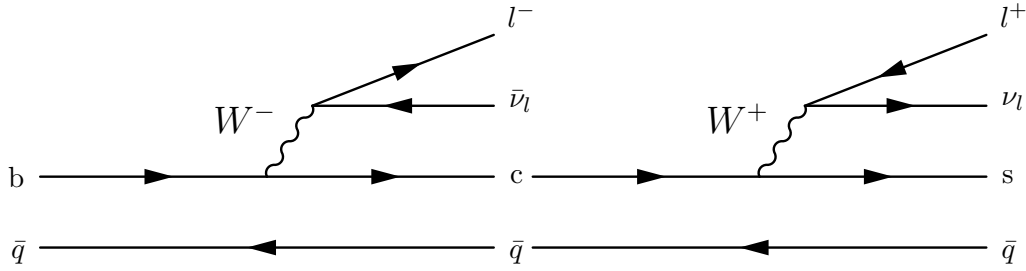


Figure 5.2: The cascade decay $b \rightarrow c \rightarrow s$

Leptons

One method to tag the flavor of the opposite side b quark is to identify the lepton coming from a semileptonic b hadron decay. The charge of the lepton is directly correlated to the flavor of the b quark. A positively charged lepton corresponds to a anti-b quark and negatively charged lepton to a b quark. The main error source apart from B oscillation is the cascade decay $b(\bar{b}) \rightarrow c(\bar{c}) \rightarrow l^{+(-)}$, which leads to a swapped tagging decision at decay time. The branching ratio for the decay $b \rightarrow cl\nu$ is of same order as for the decay $c \rightarrow sl\nu$. An example of a cascade decay is shown in figure 5.2.

Kaons

The cascade decay of the b quark is exploited in the opposite side Kaon tagging. Because the cascade decay favored by the CKM matrix ends in a s or anti-s quark, it is often accompanied by a Kaon. One can then determine the b flavor at decay time by identifying the charge of the Kaon. A K^+ corresponds to a b quark and a K^- to \bar{b} quark at production time.

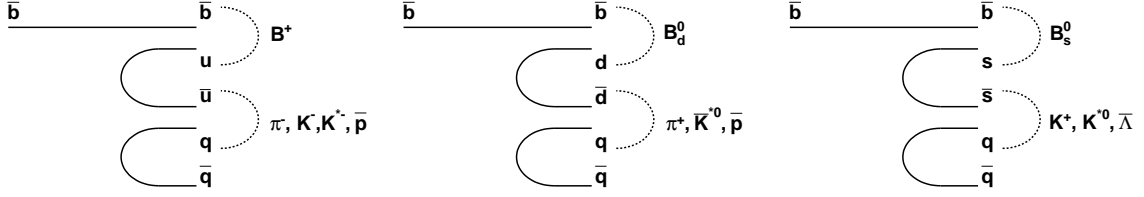


Figure 5.3: Fragmentation processes of the same side b quark

Jet Charge

Jets are produced in the process of fragmentation of quarks into hadrons. Therefore one can utilize the correlation of the jet charge to the initial quark as a tagging information. This is done by calculating a weighted sum of the charges of all tracks contained in the jet. The jet charge tagger is almost independent of the opposite side b hadron type and its decay. Its main disadvantage is the complicated task of assigning tracks to the right jet.

Vertex Charge

The method of vertex charge tagging can only be applied to charged opposite side b hadrons. To determine the b flavor one needs to reconstruct the tracks originating from the B decay vertex. The b flavor then corresponds to the sum of the charges of these tracks. A negative vertex charge indicates a b quark, a positive vertex charge a anti-b quark. The downside of this method is its bad efficiency. This is due to the fact that the reconstruction of the decay vertex of the opposite b hadron is challenging.

5.1.3 Same Side Tagging

The idea of same side tagging is to determine the flavor of the same side b quark at the production time by identifying the fragmentation tracks. In the case of a reconstructed $B_s^0(B_d^0)$ meson, one searches for a Kaon (Pion) track. The Kaon (Pion) emerges from the remaining quark of the hadronization of the B meson and an additional $u\bar{u}$ pair from the vacuum. The charge of the Kaon (Pion) then correlates to the flavor of the b, e.g. a K^- corresponds to b quark. Examples of a fragmentation processes on the same side are displayed in figure 5.3.

5.2 B Flavor Tagging at CDF II

B flavor tagging at the CDF II experiment is more challenging than in a clean e^+e^- environment. The dominant b quark production mechanism at the CDF II experiment is the production of $b\bar{b}$ pairs. In the hadronic environment there are not only the $b\bar{b}$ pairs one is interested in, but also a huge number of background tracks. These background tracks originate from fragmentation remnants or multiple interaction per beam crossing. Furthermore the $b\bar{b}$ pairs are generally not produced back to back, so that a separation into hemispheres is very difficult. In figure 5.4 this situation is represented for a typical event at the CDF II experiment.

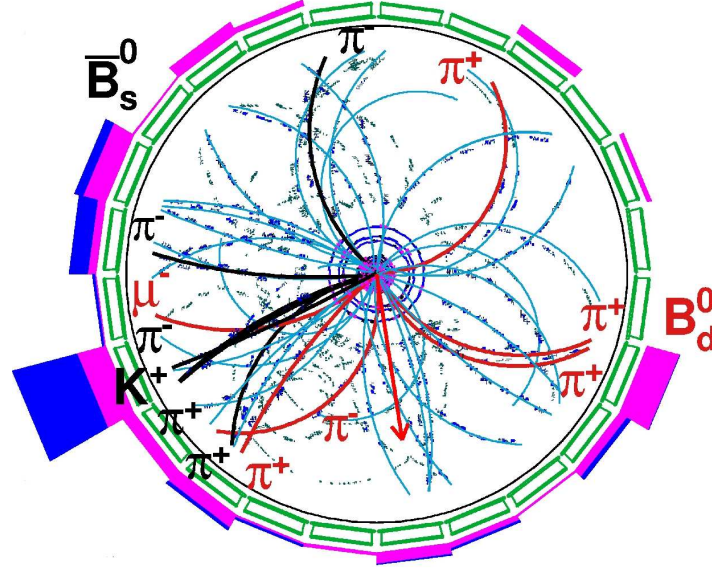


Figure 5.4: CDF II event display of a simulated $b\bar{b}$ pair, hadronizing into a \bar{B}_s^0 meson and a B_d^0 meson.

In the following the different tagging methods currently used at the CDF II experiment in the B_s oscillation analysis are pointed out and then the Neural Network B Flavor Tagger, that was recently developed by the Karlsruhe CDF B group, is introduced.

The flavor taggers currently implemented in the B_s oscillation analysis [4] are :

- on the opposite side the tagging information of kaon, lepton and jet charges are combined using an ANN to determine the opposite tag. The tagging power of the combined tagger is $T = \epsilon D^2 = 1.8\%$ [77].
- the same side tag consists of the combination of the kaon particle identification likelihood and kinematic quantities of the kaon into a tagging variable V using an ANN. The track with the largest V and close in phase space to the B_s candidate is selected as tagging track. The effectiveness of the same side tagger is $T = \epsilon D^2 = 3.7(4.8)\%$ for the hadronic (semileptonic) decay sample [10]. The different tagging powers can be explained by the differences in the composition of hadronic and semileptonic samples.
- if both same and opposite tags are available, the information is combined under the assumption that the tags are uncorrelated. The predicted dilution of the combination of the tags can be expressed by

$$\begin{array}{ll}
 \text{tag sign is equal} & \text{tag sign is opposite} \\
 D = \frac{D_1 + D_2}{1 + D_1 \cdot D_2} & D = \frac{D_1 - D_2}{1 - D_1 \cdot D_2}
 \end{array}$$

If the two tagging decisions disagree, the tagging decision with the larger dilution is taken.

5.2.1 Karlsruhe Neural Network B Flavor Tagger

The main idea of the tagger is to combine all available tagging information into a single tagging information using several neural networks. All neural networks implemented in this tagger make use of the NeuroBayes[®] package. The structure of the tagger is based upon the BSAURUS package [78] developed at the DELPHI experiment. The determination of the tagging information can be divided into two major components. At first the tagging decision is obtained on a track level basis. Afterwards this information is used to retrieve the tagging information on event level. This paragraph only gives a brief summary of the principles of the tagger. A more detailed description can be found in [11].

Flavor Tagging on Track Level

The determination of the tagging decision on the track level consists of two parts. The first part is the identification of tracks coming from the decay of the opposite b hadron and the reconstruction of the opposite side secondary vertex. The second part includes the flavor tagging for same and opposite side tracks. All tracks used for a tagging decision on track level have to pass the following precuts :

- the track is not used in the exclusive reconstruction of a B meson
- $p_t > 400$ MeV
- $d_0 < 5$ mm
- the distance Δz_0 between the current track and z_0 of all tracks originating from the same side B meson has to be less than 2 cm at the point of the closest approach of the track to the origin of the coordinate system

The tracks are subsequently used as an input of the first B track neural network. This network is trained for the identification of tracks originating from the B decay chain. The output of the first network serves as an input to the secondary vertex neural network and the second B track neural network. The secondary vertex neural network only accepts the four tracks with the highest neural network output from the first B track neural network and combines them into six vertex candidates. The information for the best vertex candidate is passed on to the next network. The second B track neural network has the same purpose as the first B track neural network, but it is trained with the additional information of the first B track network and the opposite side secondary vertex.

The next level of the tagging algorithm is the flavor tagging part on track level. The first network at this stage is the 'Lepton from B' neural network, which only uses tracks passing precuts on the lepton identification variables. This network decides if a lepton comes from a B decay chain. If the probability for a lepton passes a certain cut, the track is used in the lepton flavor neural network. The task of the lepton flavor network is to identify the flavor of the b quark using the correlation of the charge of track to the b flavor and to separate leptons of the B decay chain from the D decay chain. Therefore the network mainly uses lepton identification quantities and track information relating to the secondary vertex.

The last stage of the flavor tagging on track level is the main flavor network. It is the core component of the track level tagging and combines the output of the second B track network, the secondary vertex network, the 'lepton from B' network and the lepton flavor network into a single flavor probability of the track. An overview of the different networks and the basic workflow of the tagging procedure on track level is shown in figure 5.5.

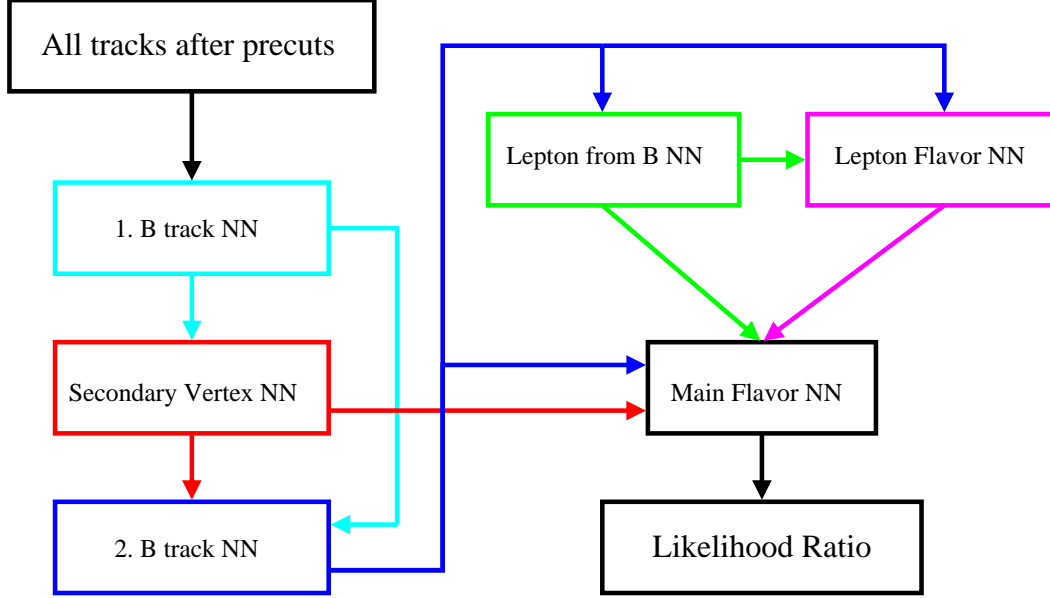


Figure 5.5: Workflow of the flavor tagger on track level.

Flavor Tagging on Event Level

To obtain a tagging decision on event level, one has to combine the information provided by the main flavor network on the track level into a single probability. This is done in a likelihood ratio under the assumption that the probabilities of each track are independent. In addition to this likelihood ratio, more quantities for tagging on event level are available. The opposite vertex charge is calculated as a weighted sum of the charges of all tracks in an event with the B track probabilities used as weights. Another quantity is the tagging decision provided by the jet charge tagger. As further additional information the two best leptons, the best same side kaon and the best opposite side kaon are used. The quantities mentioned above and the likelihood ratio form the input variables for the combined tag network as displayed in figure 5.6. The output of the combined tag network then gives the final tagging probability for an event. The performance of this tagging procedure is estimated on Monte Carlo samples. The result [11] for the tagging power is

$$T = \epsilon D^2 = 9.31\% \quad (5.3)$$

This improves the tagging power compared to existing taggers by a factor of two. It is the ambition of this thesis to check this large enhancement on data. This is done by measuring the amplitude and the error on the amplitude in the mixing analysis as described in chapter 8 of this thesis.

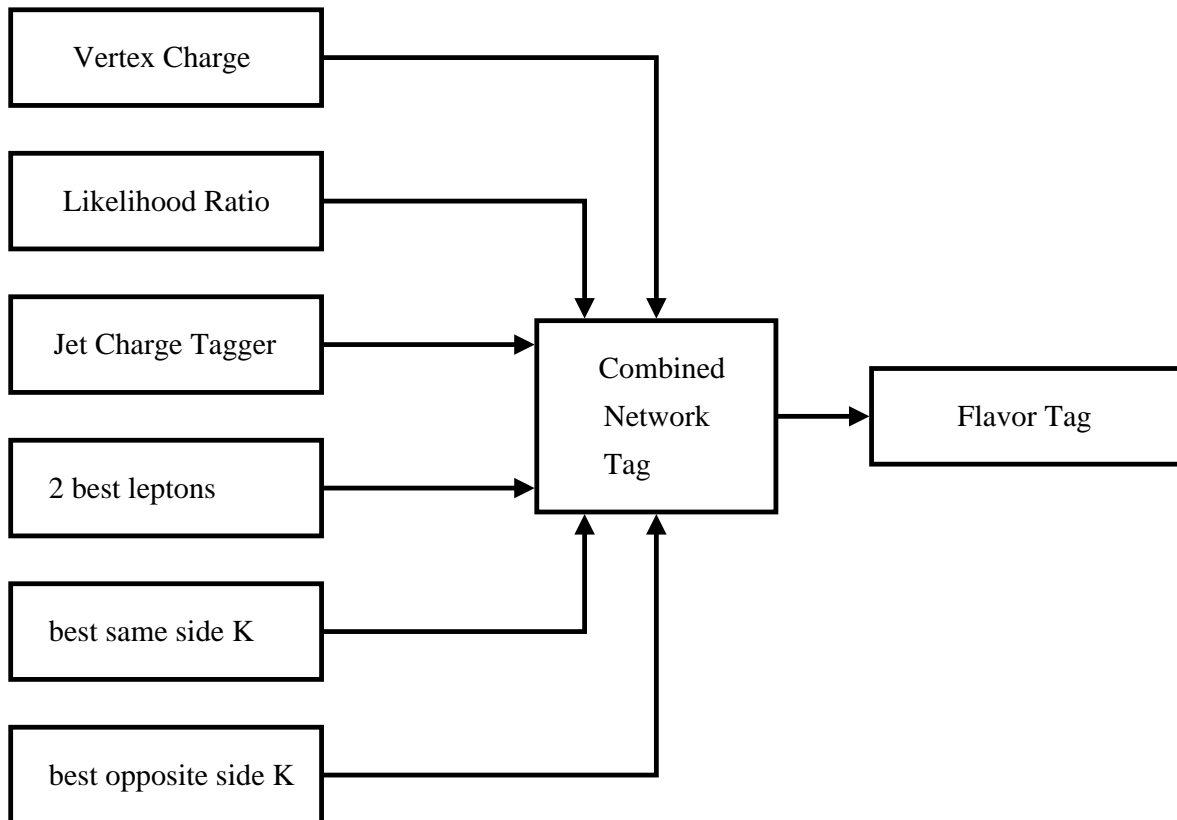


Figure 5.6: Flowchart of the flavor tagger on event level.

Chapter 6

Exclusive B_s Mesons Selection

6.1 Motivation

The importance of a good selection of the exclusively reconstructed B_s meson is clearly apparent from the significance of the Δm_s measurement, see equation 4.4. This chapter discusses the improvement of the $\frac{N_s}{\sqrt{N_s+N_B}}$ term of the significance by introducing neural networks for the selection of signal events. The use of neural networks instead of cut based methods has several benefits. The cut based approach tries to optimize the signal significance by varying cuts on one single variable while all other are kept fixed. The cut based selection corresponds to a rectangular selection in the space spanned by the variables. However a neural network is able to combine all the information of the variables in an optimal way, because its selection forms a more complex shape in the space of the variables and thereby considers the correlations between the variables. The aim of this chapter is to show the gain in significance of the Δm_s measurement using neural network based selections instead of cut based selections.

6.2 Optimizing the Selection of Exclusive B_s Mesons

The optimization process for the selection of exclusively reconstructed B_s mesons requires a profound understanding of the properties of the examined decay channels. This section is divided into three major parts.

Decay Mode	Branching Ratio
$B_s \rightarrow D_s \pi$	$(3.8 \pm 0.3(stat.) \pm 1.3(syst.)) \times 10^{-3}$ [79]
$B_s \rightarrow D_s 3\pi$	$(8.4 \pm 0.8(stat.) \pm 3.2(syst.)) \times 10^{-3}$ [79]
$D_s \rightarrow \phi \pi$	$2.16 \pm 0.28\%$
$\phi \rightarrow KK$	$49.2 \pm 0.6\%$
$D_s \rightarrow K^* K$	$2.5 \pm 0.5\%$
$K^* \rightarrow K \pi$	$\sim 100\%$
$D_s \rightarrow 3\pi$	$1.22 \pm 0.23\%$

Table 6.1: Summary of the hadronic decay modes studied in the analysis. Unless specified the world averages of the branching fractions are given [16].

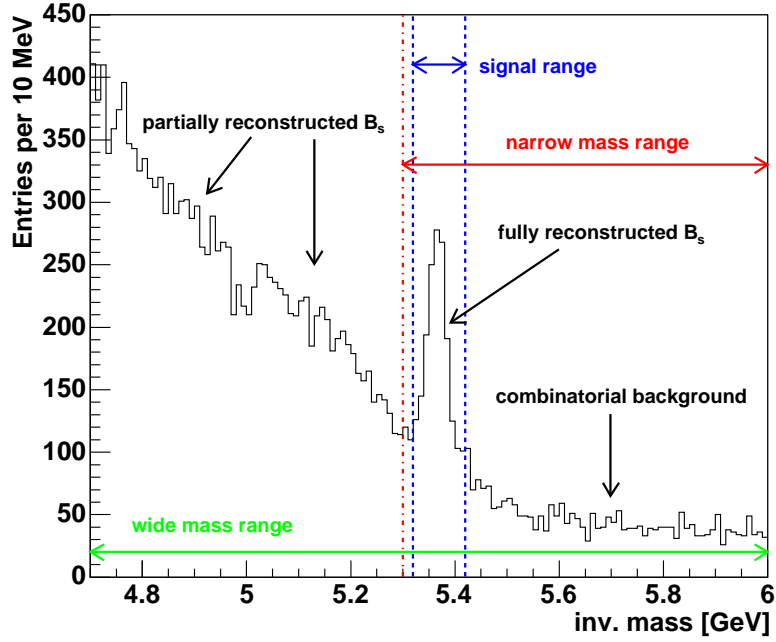


Figure 6.1: Example of a typical mass spectrum for the decay of a B_s meson. Here the decay $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$ is shown.

The first part describes the B_s decay modes and the general properties of the invariant mass spectra. The second part explains the training of the neural networks. Finally the unbinned and binned maximum likelihood fits to the mass spectra are performed and the results are presented in the third part. This analysis only considers the hadronic decay modes of the B_s meson listed in table 6.1. The three pions of the $D_s3\pi$ decay mode originate in almost all cases from an intermediate $a_1(1260)$. The $a_1(1260)$ is a very broad resonance and decays over $a_1(1260) \rightarrow \rho\pi \rightarrow \pi\pi\pi$. The three pions coming from a D_s have a different decay topology. They are final states of the following decays through intermediate meson states : $D_s \rightarrow \rho\pi$, $f_0(980)\pi$, $f_2(1270)\pi$, $f'_0(1370)\pi \rightarrow \pi\pi\pi$

6.2.1 The Mass Spectra of the B_s Decay Channels

The mass spectra used for the binned fits of the studied B_s decay channels range from 4.7 to 6.0 GeV, because the invariant mass of the B_s is 5.369 GeV, and all information relevant for the background parameterization is contained in this mass window. In this analysis the mass range of [4.7, 6.0] GeV is called wide mass range. The narrow mass range used for all the unbinned fits is [5.3, 6.0] GeV. The signal significance is calculated in the mass range of [5.32, 5.42] GeV corresponding to an approximately 2.5σ window. In this paragraph only the common features of the mass spectra of all decay channels are outlined. An example for the shape of a mass spectrum is displayed in figure 6.1. The details of the fit functions used to model the mass spectra and the exact composition of the mass spectra for each decay mode are described in paragraph 6.2.3.

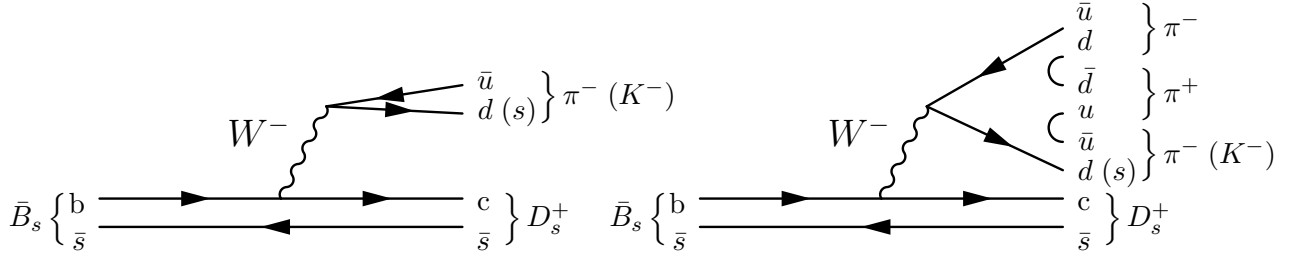


Figure 6.2: Feynman diagrams for the tree level contribution of the decays $B_s \rightarrow D_s(3)\pi$

Signal Contributions

The B_s meson decay into $D_s(3)\pi$ has two signal contributions, the cabibbo favored decay $B_s \rightarrow D_s(3)\pi$ and the cabibbo suppressed decay $B_s \rightarrow D_s K(\pi\pi)$. The ratio of the cabibbo favored to the cabibbo suppressed decays is expected to be $\frac{|V_{ud}|^2}{|V_{us}|^2} \sim 20$. The events of the cabibbo suppressed decay are counted as signal events for all studied decay modes. The tree level Feynman diagrams for both signal contributions without the intermediate meson states are displayed in figure 6.2. The intermediate meson states of the cabibbo suppressed decay $B_s \rightarrow D_s K\pi\pi$ are not well understood. One assumes the following intermediate states for the Monte Carlo production of cabibbo suppressed decays :

- $B_s \rightarrow D_s \rho K$
- $B_s \rightarrow D_s K_1$
- $B_s \rightarrow D_s D$.

Background Contributions

The background contributions to the mass spectra can be divided into three categories :

- The first category of background events consists of fully reconstructed decays of hadrons mimicking the decay topology of the studied B_s decays. In these cases a wrong particle hypothesis has been assigned to one particle in the final state. The particles contributing to the background in the channels used for this analysis are the Λ_b and the B^0 :

- $\Lambda_b \rightarrow \Lambda_c(3)\pi$
 $\Lambda_c \rightarrow pK\pi$
 $\Lambda_c \rightarrow p\pi\pi$
- $B^0 \rightarrow D(3)\pi$
 $D \rightarrow K\pi\pi$

The B^0 and Λ_b contributions are of great importance to the unbinned maximum likelihood fits, because both have similar physical properties as the B_s . Furthermore this assures that the oscillation of B^0 mesons is accounted for the unbinned maximum likelihood function of the mixing fit.

- Events originating from real B_s mesons where neutral particles are not reconstructed are called partially reconstructed events. In this analysis the partially reconstructed decays are subdivided into the following three classes :

$$\begin{aligned}
- B_s &\rightarrow D_s^*(3)\pi \\
&\quad D_s^* \rightarrow D_s\gamma \\
- B_s &\rightarrow D_s \rho (\pi\pi) \\
&\quad \rho \rightarrow \pi\pi_0 \\
- B_s &\rightarrow D_s X
\end{aligned}$$

The X denotes all possible combinations of particles with the D_s meson coming from the decays of the B_s meson, except the $D_s^*\pi$ and the $D_s\rho$ decay modes. The semileptonic decays $D_sl\nu$ are also contained in the D_sX category. Due to the missing particles the invariant mass of these decays is smaller than the B_s mass. Therefore the partially reconstructed decays only contribute to lower part of the wide mass range. They are neglected in the unbinned maximum likelihood fits, which are only performed in the narrow mass range.

- Background events from track combinations that pass all requirements for a B_s but do not originate from a B_s meson, are called combinatorial background. The combinatorial background is the largest background contribution to the mass spectra in all studied mass ranges. This contribution can be strongly reduced by the use of a neural network selection or a good cut based selection of the events.

6.2.2 Training and Application of Neural Networks in the Selection of Exclusively Reconstructed B_s Mesons

General Strategy

The strategy used for the training of the neural networks is the same for all studied B_s decay channels. Only the channel $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi\pi$ possesses a peculiarity, namely two additional boost networks. The method of a boost training is explained in [6]. To separate signal from background events in the exclusive decay modes of the B_s mode, one uses two input classes for the training. The first input class consists of signal events, generated with a Monte Carlo event generator. The second input class is formed of background events obtained from a data sample. This is due to the lack of a Monte Carlo simulation that is able to model the combinatorial background correctly in the complex hadronic environment. The background events are taken from the upper sideband of the mass spectrum. Therefore one can assure that the background events only consists of combinatorial background. Further one assumes that the background of the upper sideband holds as a good description of the background in the signal region.

After the preparation of data and Monte Carlo samples, one has to determine reasonable input variables for the neural network. The neural network then is trained with these input variables and the training patterns provided by the data and Monte Carlo samples in the 'Teacher'. The output of this training is written to an expert file, called expertise. This expertise is used by the 'Expert' to classify a given data sample by calculating the neural network output.

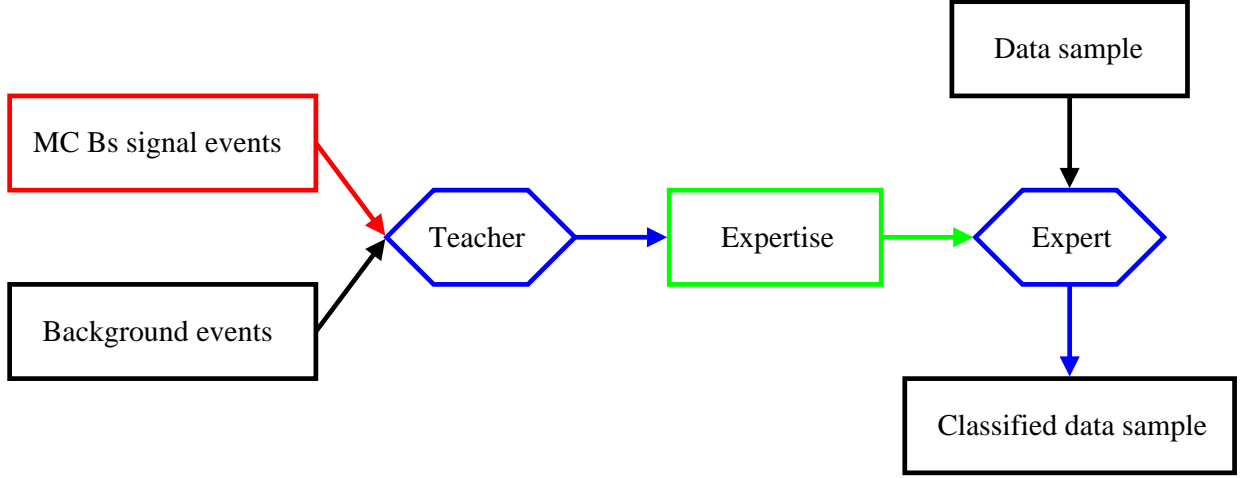


Figure 6.3: Schematical representation of the classification procedure using NeuroBayes[®].

The cut on the neural network output determines a working point for the final selection of B_s mesons. In this analysis one is interested in the working point that maximizes the significance $\mathcal{S} = \frac{S}{\sqrt{S+B}}$ of the studied decay channel. In figure 6.3 a schematic overview of the strategy for the application of neural networks to the exclusive reconstruction of B_s mesons is displayed.

Preparing the Data and Monte Carlo Files for Training

The Monte Carlo samples for the signal events are generated by BGEN, except the samples for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$. This decay uses PYTHIA as an event generator. To mimic the effects of trigger prescaling in the Monte Carlo sample, run-averaged prescale factors for the different trigger paths are used. For all data and MC samples throughout this analysis a set of cuts independent of the particular final state is applied :

- Number of COT Stereo/Axial Hits $\geq 10/10$
- Number of Silicon Stereo + Silicon Axial Hits ≥ 3
- all tracks have $p_t > 0.35$ GeV
- B_s meson passes one of the B_LOWPT, B_CHARM or the B_HIGHPT trigger paths. In the case of Monte Carlo events only Level 2 trigger paths are checked due to the lack of a Level 3 simulation.

To further reduce the size of data samples used for the training of neural networks a set of preselection cuts is applied. These cuts aim at eliminating a large fraction of background events, without cutting away a significant number of signal events. In order to train a neural network these cuts have to be applied to the Monte Carlo signal events as well. The preselection cuts are summarized in table 6.2. The background events of the upper sideband are selected from the data samples after the the preselection cuts have been applied. The choice of the mass range for the upper sideband in the interval $[5.55, 6.55]$ GeV is arbitrary. A detailed study on the influence of the upper sideband selection to the output of the network

has been performed in [5] and no significant effect has been observed. The mass ranges of the upper sideband are listed in table 6.3. In figure 6.4 an example of a signal and background sample for the training of the neural network is shown.

preselection cuts on	$B_s \rightarrow D_s \pi$ modes			$B_s \rightarrow D_s 3\pi$ modes		
	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$
$\chi_{r\phi}^2(B_s)$	< 20	< 20	< 20	< 20	< 20	< 20
$\chi_{r\phi}^2(D_s)$	< 20	< 20	< 20	< 20	< 20	< 20
$L_{xy}/\sigma_{L_{xy}}(B_s)$	> 2	> 2	> 2	> 2	> 2	> 6
$\sigma_{L_{xy}}(B_s)$ [cm]	-	-	-	< 0.015	< 0.015	-
$L_{xy}/\sigma_{L_{xy}}(D_s)$	> 2	> 2	-	> 2	> 2	> 6
$ d_0(B_s) $ [cm]	< 0.02	< 0.02	< 0.009	< 0.01	< 0.012	< 0.015
$p_t(B_s)$ [GeV]	> 5.5	> 5.5	> 5.5	> 5.5	> 5.0	> 4.0
$p_t(\pi_{B_s})$ [GeV]	> 1.0	-	> 0.4	-	-	-
$\min p_t$ [GeV]	> 0.35	> 0.4	> 0.4	> 0.35	> 0.35	> 0.35
$L_{xy}(B_s \leftarrow D_s)$ [cm]	-	-	[-4, 4]	-	> -0.05	> -0.1
$ \Delta m(K\pi\pi, D^-) $ [GeV]	-	> 0.024	-	-	> 0.024	-
$q(K_{K^*}) \cdot q(K_{D_s})$	-	< 0	-	-	< 0	-
inv. mass $m_{3\pi_{B_s}}$ [GeV]	-	-	-	-	-	< 2.5

Table 6.2: Summary of preselection cuts for all $B_s \rightarrow D_s(3)\pi$ decay modes.

	$B_s \rightarrow D_s \pi$ modes			$B_s \rightarrow D_s 3\pi$ modes		
	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$
mass range [GeV]	5.6 – 5.9	5.55 – 5.85	5.6 – 6.55	5.6 – 6.55	5.6 – 5.8	5.6 – 5.9

Table 6.3: Mass ranges for the upper sideband used for the network trainings.

Selection of Input Variables for the Neural Network

The choice of variables for the training is not completely random. The variables used for the training are subject to several conditions. They have to discriminate between signal and background and their description has to be consistent in the data and Monte Carlo samples. The 'Teacher' implemented in the NeuroBayes[®] software package helps in the selection of the variables. In general one selects variables already used in a cut based analysis or variables that show a good discrimination between signal and background. The output of the training by the 'Teacher' contains a list of the correlations of the variables to the target and among each other. Furthermore the significance of each variable is included in this list. The next step in the selection of variables is to remove highly correlated variables or variables that only add a small amount of new information to the neural network. The general rule is to keep the number of training variables as low as possible to improve the generalization capabilities of the neural network. In case of the optimization of the selection of exclusively reconstructed B_s mesons one has to check if the neural network learns to separate signal from background events or only distinguishes the events by their invariant masses.

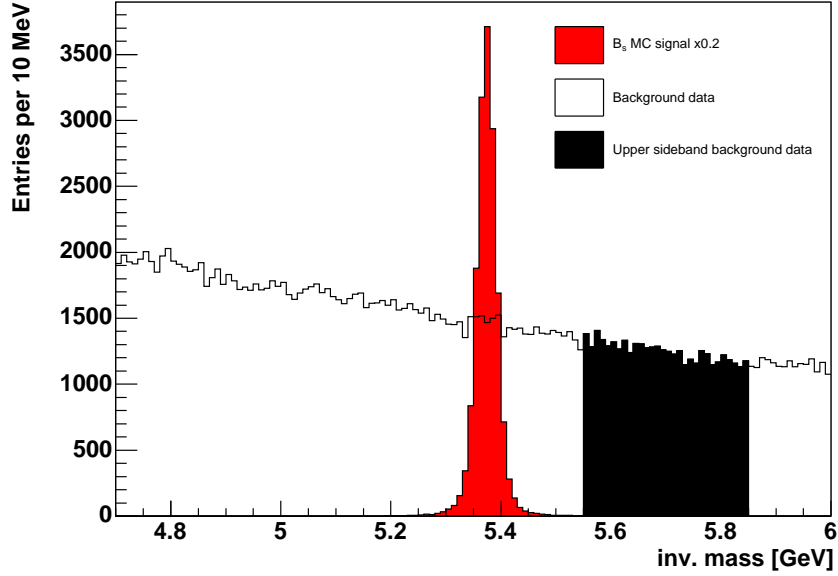


Figure 6.4: Example of a data and a MC sample used for the training of the neural network in the channel $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$. For a better illustration the Monte Carlo signal is scaled down to a fifth of the original distribution.

A detailed study of variables indicating a mass dependence and their influence to the training of neural networks can be found in the thesis [5]. In table 6.4 a summary of variables used in the training of neural networks for all channels is given. The meaning of each variable is explained in Appendix A. In Appendix B a compilation of the training results is presented. The channels $B_s \rightarrow (3)\pi$, $D_s \rightarrow K^*K$ and $B_s \rightarrow (3)\pi$, $D_s \rightarrow 3\pi$ and their particular characteristics are discussed in the following. For all other channels one can refer to [5,6] for detailed studies of the network variables and the properties of the corresponding decays.

The decays $B_s \rightarrow D_s(3)\pi$, $D_s \rightarrow K^*K$ and $B_s \rightarrow D_s3\pi$, $D_s \rightarrow 3\pi$

The decays of the one pion and the three pion channel are treated the same way for all variables regarding the $D_s \rightarrow K^*K$ decay. In these channels the B_s signal is mainly polluted with B^0 events. This is due to the resemblance of final states of the decay $B^0 \rightarrow D(3)\pi$ and $B_s \rightarrow D_s(3)\pi$. The stable decay products of these modes are $K\pi\pi\pi$ for the B^0 and $KK\pi\pi$ for the B_s . Therefore the only difference during the reconstruction of a B_s is the wrong identification of a pion as a kaon. To reduce the number of B^0 , a cut on the mass difference of the D_s and D is performed. A further possibility to decrease the number of B^0 is the use of particle identification variables for the neural network training. This is yet not implemented in this analysis, because at the time of these studies the variable description in the Monte Carlo samples disagreed with the data. The remaining fraction of the B^0 events in the signal region is $< 5\%$ (one pion mode) and $< 10\%$ (three pion mode) depending on the neural network cut.

Beside the commonly used variables, like χ^2 , L_{xy} , p_t , d_0 , two more variables are interesting in the case of the K^* decay : the helicity angle and the K^* invariant mass. The helicity angle θ_{hel} of the kaon coming from the K^0 decay is the angle measured in the K^* rest frame between the direction of the kaon and the direction of the D_s meson. The helicity distributions for Monte Carlo signal and combinatorial background events for the decay $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$ are shown in figure 6.5. The angular distribution for the signal follows a $\cos^2(\theta_{hel})$ function, whereas the background events are nearly uniformly distributed. The \cos^2 shape of the angular distribution can be explained by the spin composition of the D_s decay :

$$\begin{array}{rcccl}
 & D_s & \rightarrow & K^* & K & , & K^* & \rightarrow & K & \pi \\
 J & = & 0 & \rightarrow & 1 & 0 & , & 1 & \rightarrow & 0 & 0
 \end{array}$$

The invariant mass of the K^* is as well a good variable for the separation of signal and background, illustrated in figure 6.6. The absolute value of the deviation of the measured invariant mass and the world average value, taken from [16], serves as an input variable for the neural networks. Further variables make use of the invariant masses of the pions originating from the B_s and their angular correlations, which are also common variables to the decay $B_s \rightarrow D_s3\pi$, $D_s \rightarrow 3\pi$.

In the $B_s \rightarrow D_s3\pi$, $D_s \rightarrow 3\pi$ decay mode the combinatorial background is very high due to the huge number of possible combinations of the six tracks of the final state pions. To decrease the number of background events one can exploit the information provided by the a_1 resonance and the various intermediate meson states of the $D_s \rightarrow 3\pi$ decay. Therefore the invariant mass distributions of the three pion vertex and the invariant masses of different pion combinations originating from the three pion vertex or D_s are used as input variables for the neural network. In figure 6.7 the invariant mass distribution of the three pion vertex is shown. The invariant mass $m_{\pi_{D_s}^2, \pi_{D_s}^3}$ of the four momentum $p_{\pi_{D_s}^2, \pi_{D_s}^3} = (p_{\pi_{D_s}^2} + p_{\pi_{D_s}^3})$ serves as a good example for the importance of the intermediate meson states of the D_s decay. In this case the main signal contributions come from the $f_0(980)$ meson as represented in figure 6.9. Thereby one has to mention that it was recently realized that a mass of 1000 MeV was wrongly assigned to the $f_0(980)$ meson in the generation of the Monte Carlo sample. An additional interesting variable is the angle between the momentum of the three pion vertex and the plane spanned by the momenta of two pions coming from the B_s decay. This angle is further denoted by $\theta(\vec{n}_{(\pi_{B_s}^1, \pi_{B_s}^2)}, \vec{p}(3\pi))$ and an example of its distribution for signal and background events is given in figure 6.8 .

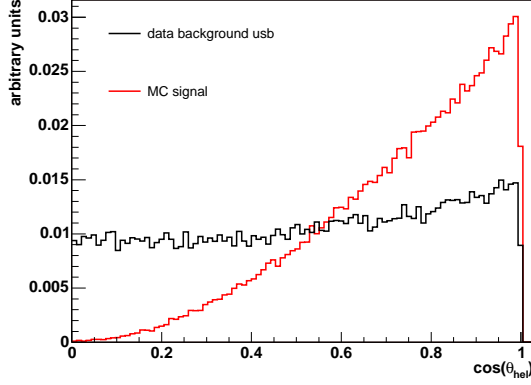


Figure 6.5: Helicity angle of the kaon for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ for MC signal and upper sideband background.

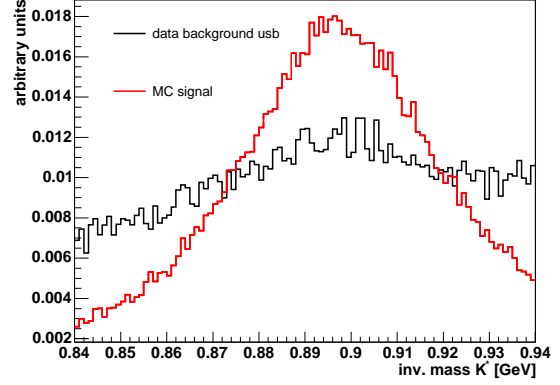


Figure 6.6: Invariant mass of the K^* for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ for MC signal and upper sideband background.

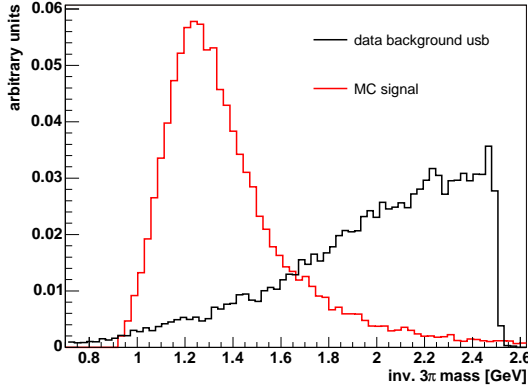


Figure 6.7: Invariant mass of the 3π vertex for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$ for MC signal and upper sideband background.

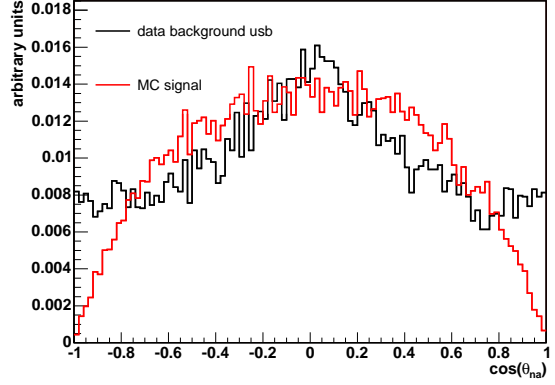


Figure 6.8: Distribution of $\theta(\vec{n}_{(\pi_{B_s}^1, \pi_{B_s}^2)}, \vec{p}(3\pi))$ for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$ for MC signal and upper sideband background.

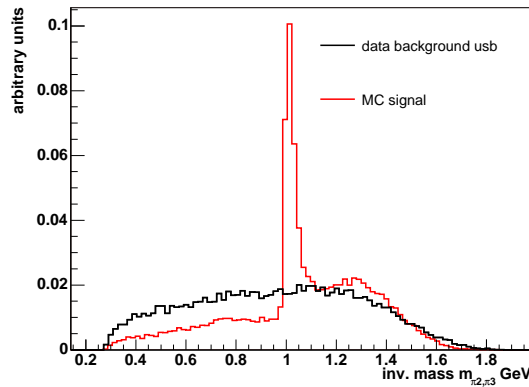


Figure 6.9: Invariant mass distribution of $m_{\pi_{D_s}^2, \pi_{D_s}^3}$ for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$ for MC signal and upper sideband background.

variable name	$B_s \rightarrow D_s \pi$ modes			$B_s \rightarrow D_s 3\pi$ modes		
	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$
$L_{xy}/\sigma_{L_{xy}}(B_s)$	✓	✓	✓	✓	✓	✓
$\sigma_{L_{xy}}(B_s)$				✓	✓	✓
$L_{xy}/\sigma_{L_{xy}}(D_s)$	✓	✓		✓		
$L_{xy}(B_s \leftarrow D_s)$	✓	✓	✓	✓	✓	✓
$\chi_{3D}^2(B_s)$	✓	✓			✓	✓
$\chi_{3D}^2(D_s)$	✓	✓				✓
$\chi_{T\phi}^2(B_s)$		✓	✓	✓	✓	✓
$\chi_{T\phi}^2(D_s)$			✓	✓	✓	✓
$p_t(B_s)$			✓	✓		
$p_t(D_s)$		✓			✓	✓
$p_t(\pi_{B_s})$	✓	✓	✓			
$p_t(\pi_{B_s}^1)$				✓	✓	
$p_t(\pi_{B_s}^2)$				✓		
$p_t(\pi_{B_s}^3)$				✓		
$p_t(K_\phi^1)$				✓		
$p_t(K_\phi^2)$				✓		
$\min p_t$	✓	✓			✓	✓
$ d_0(B_s) $	✓	✓	✓	✓	✓	✓
$ d_0(D_s) $	✓				✓	✓
lts. $d_0/\sigma_{d_0}(K_\phi^1)$	✓					
lts. $d_0/\sigma_{d_0}(\pi_{K^*})$		✓				
lts. $d_0/\sigma_{d_0}(\pi_{B_s})$			✓			
lts. $d_0/\sigma_{d_0}(\pi_{D_s}^n)$			✓			
$\min d_0/\sigma_{d_0}$		✓			✓	✓
\min lts. d_0/σ_{d_0}					✓	✓
$\min d_0(\pi_{B_s}^1, \pi_{B_s}^2, \pi_{B_s}^3)$				✓		
$\max d_0(\pi_{B_s}^1, \pi_{B_s}^2, \pi_{B_s}^3)$				✓		
helicity angle K_ϕ^1	✓					
helicity angle K_{K^*}		✓			✓	
CMS angle π_{B_s}			✓			
$\theta(\vec{n}(\pi_{B_s}^1, \pi_{B_s}^2), \vec{p}(3\pi))$				✓	✓	✓
$m_{\phi \rightarrow KK}$	✓			✓		
$ m_{K^*} - m_{K^*(PDG)} $		✓			✓	
inv. mass $m_{\pi_{D_s}^1 \pi_{D_s}^2}^2$			✓			
inv. mass $m_{\pi_{D_s}^1 \pi_{D_s}^3}^2$			✓			
inv. mass $m_{\pi_{D_s}^2 \pi_{D_s}^3}^2$			✓			
inv. mass $m_{\pi_{D_s}^1 \pi_{D_s}^2}$						✓
inv. mass $m_{\pi_{D_s}^1 \pi_{D_s}^3}$						✓
inv. mass $m_{\pi_{D_s}^2 \pi_{D_s}^3}$						✓
inv. mass $m_{\pi_{B_s}^1 \pi_{B_s}^2}$				✓	✓	✓
inv. mass $m_{\pi_{B_s}^1 \pi_{B_s}^3}$						✓
inv. mass $m_{\pi_{B_s}^2 \pi_{B_s}^3}$						✓
$\min(m_{\pi_{B_s}^{12}}, m_{\pi_{B_s}^{23}})$				✓	✓	✓
$\max(m_{\pi_{B_s}^{12}}, m_{\pi_{B_s}^{23}})$				✓	✓	✓
$\min(m_{\pi_{D_s}^{12}}, m_{\pi_{D_s}^{23}})$						✓
$\max(m_{\pi_{D_s}^{12}}, m_{\pi_{D_s}^{23}})$						✓
inv. mass $m_{3\pi_{B_s}}$					✓	✓
m_{diff}				✓		
$\Delta R(D_s, \pi_{B_s})$			✓			

Table 6.4: Variables used in the neural network trainings

6.2.3 Fitting the Mass Spectra

To evaluate the performance of the neural networks and to determine the parameters for the mixing likelihood function, one fits the mass spectra using the following probability density function. The function is designed to be used as a general function valid for all studied $B_s \rightarrow D_s(3)\pi$ decay modes. The probability density function in the mass space can be written as

$$P(m) = N \cdot ((1 - f_b) \cdot P_S(m) + f_b \cdot P_B(B)) \quad (6.1)$$

with $f_b = \frac{N_B}{N_S + N_B}$ being the background fraction in the fitted mass range and N the overall normalization.

The B_s signal is modeled by a single Gaussian function. The shape of the cabibbo suppressed decay is described by the convolution of a Gaussian and an exponential function.

$$P_S(m) = (1 - f_{Cabibbo}) \cdot P_{Signal} + f_{Cabibbo} \cdot P_{Cabibbo} \quad (6.2)$$

$$P_{Signal} = \mathcal{G}(m, \mu_{Signal}, \sigma_{Signal}) \quad (6.3)$$

$$P_{Cabibbo} = N_{Cabibbo} \cdot \left(e^{-\frac{m}{\tau_{Cabibbo}}} \otimes \mathcal{G}(m, \mu_{Cabibbo}, \sigma_{Cabibbo}) \right) \quad (6.4)$$

where $f_{Cabibbo}$ denotes the fraction of the cabibbo suppressed decay and $N_{Cabibbo}$ is the normalization.

The background probability density function can be expressed by the composition of the combinatorial background, the partially reconstructed decays and the B^0 and Λ_b contributions.

$$P_B(m) = f_{comb} \cdot P_{comb} + f_{part} \cdot (1 - f_{comb}) \cdot P_{part} + (1 - f_{part}) \cdot (1 - f_{comb}) \cdot (f_\Lambda P_{\Lambda_b} + (1 - f_\Lambda) \cdot P_{B^0}) \quad (6.5)$$

with f_{comb} being the fraction of the combinatorial background and f_{part} the fraction of partially reconstructed decays. $f_\Lambda = \frac{N_{\Lambda_b}}{N_{\Lambda_b} + N_{B^0}}$ is the ratio of Λ_b decays to B^0 decays.

The combinatorial background is modeled using the sum of an exponential and a constant offset.

$$P_{comb} = N_{comb} \cdot \left((1 - f_{const}) \cdot e^{-\lambda(m - m_c)} + f_{const} \right) \quad (6.6)$$

here f_{const} denotes the constant offset and $m_c = 0.5 \cdot (m_{high} - m_{low})$, where m_{high} and m_{low} are the upper and lower bounds of the mass range of the fit. N_{comb} normalizes the function.

The partially reconstructed decays are fitted by the sum of a Gaussian function and a linear function.

$$\begin{aligned} P_{part} &= f_{gaus} \cdot P_{gaus} + (1 - f_{gaus}) \cdot P_{lin} \\ &= f_{gaus} \cdot \mathcal{G}(m, \mu, \sigma) + (1 - f_{gaus}) \cdot \left[\frac{2N_{lin}}{(c - m_{low})^2} \cdot (c - m)\theta(c - m) \right] \end{aligned} \quad (6.7)$$

c denotes the cutoff parameter for the linear function. f_{gaus} describes the fraction of the Gaussian function to the linear function in the partially reconstructed decays. N_{lin} normalizes the linear part of the function to one.

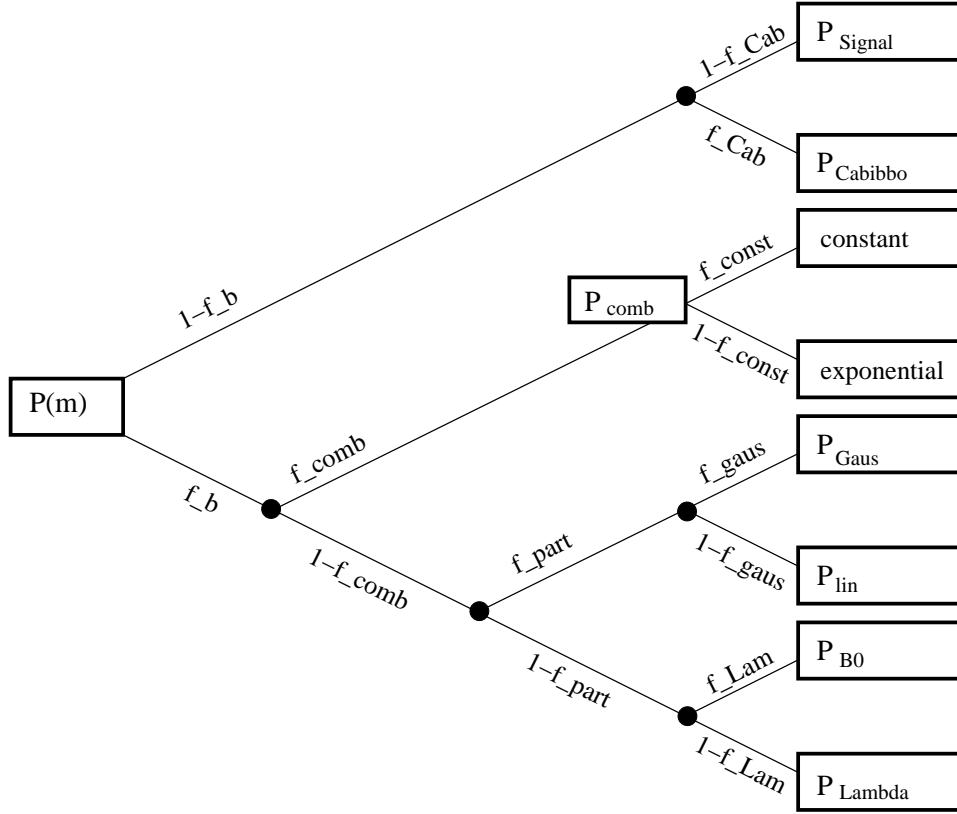


Figure 6.10: Binary tree of the general mass fit function

The shape of the physics background contribution of the Λ_b and the B^0 is described by the same function, the sum of a Gaussian function and the convolution of a Gaussian and an exponential function. It is a purely phenomenological description and the parameters have no physical meaning.

$$P_{\Lambda_b} = N_{\Lambda_b} \left(e^{-\frac{m}{\tau_{\Lambda_b}}} \otimes \mathcal{G}(m, \mu_{1,\Lambda_b}, \sigma_{1,\Lambda_b}) + \mathcal{G}(m, \mu_{2,\Lambda_b}, \sigma_{2,\Lambda_b}) \right) \quad (6.8)$$

$$P_{B^0} = N_{B^0} \cdot \left(e^{-\frac{m}{\tau_{B^0}}} \otimes \mathcal{G}(m, \mu_{1,B^0}, \sigma_{1,B^0}) + \mathcal{G}(m, \mu_{2,B^0}, \sigma_{2,B^0}) \right) \quad (6.9)$$

where N_{B^0} and N_{Λ_b} are the normalizations.

In figure 6.10 the binary tree structure of the general function is illustrated to show the relations of signal and background fractions and the corresponding probability density functions.

Before the fits are performed, one has to obtain some of the fractions and parameters from Monte Carlo templates. This is due to the low statistics of the studied exclusive B_s decay modes and therefore the simultaneous determination of these parameters by a single fit is unfeasible. The parameters and fractions obtained by the fits to the Monte Carlo templates are fixed in all further fits, except in case of the $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$ and $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$ decay modes. In these modes the description of the partially reconstructed decays in the Monte Carlo samples is not sufficient to describe the data. Therefore the parameters of the partially reconstructed decays are used as starting values and are varied within specified limits. A complete compilation of the Monte Carlo templates used for this purpose can be found in Appendix C. Because no simulation for the combinatorial background exists, the fraction of combinatorial background events and the parameters for the background shape are directly obtained in the fit.

For the final mixing fit one needs an optimal selection of exclusively reconstructed B_s mesons. Therefore the working point of the neural network, characterized by the maximum significance, is determined by a sequence of cuts on the neural network output. For each cut on the neural network output the mass spectrum is fitted in the wide mass range with the corresponding parameters and fractions received from the fits to the Monte Carlo templates. In figure 6.11 the significance as a function of the cut on the output of the neural network is displayed for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$. The corresponding result of the binned fit at the selected working point is represented in figure 6.12 .

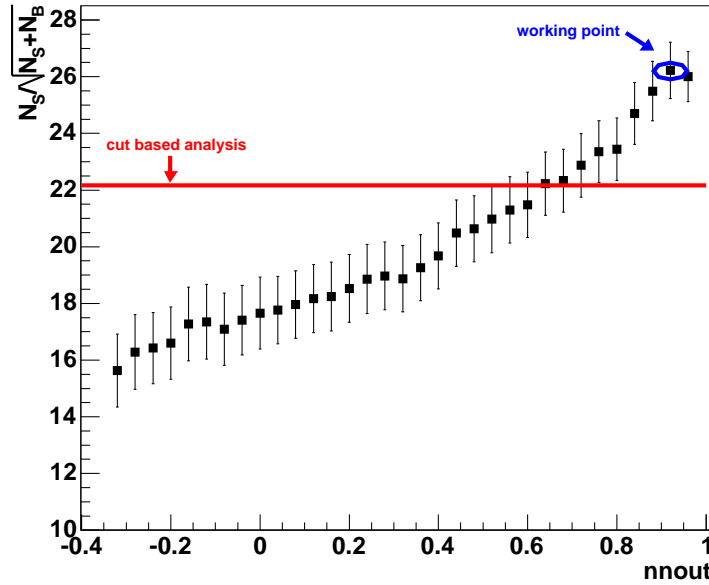


Figure 6.11: Significance as a function of the cut on the network output for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

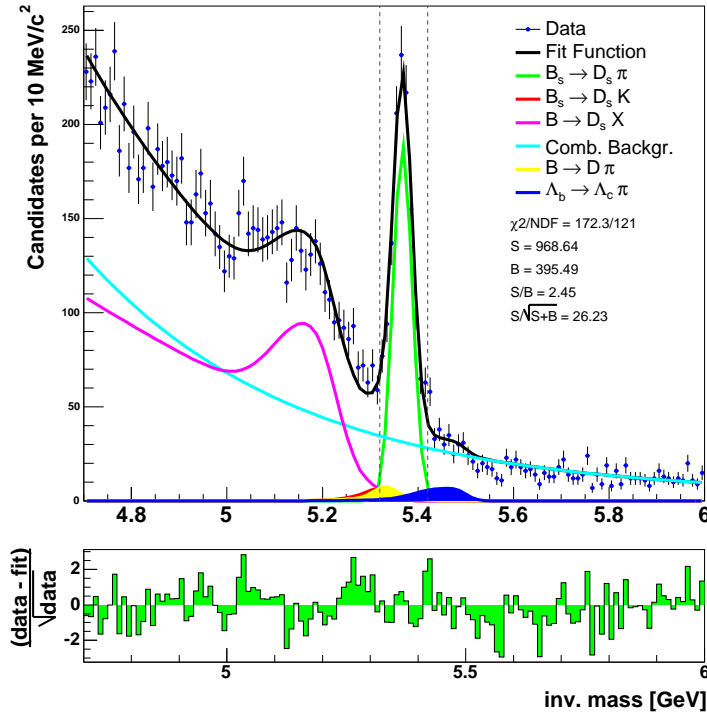


Figure 6.12: Invariant mass spectrum for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ and the binned fit result for a cut on the neural network output, $nnout > 0.92$.

To evaluate the performance of the neural networks, the results of the binned fits for the neural network selections are compared to the results of a cut based analysis. The cuts for the cut based analysis are taken from [80] and summarized in table 6.7. In tables 6.5 and 6.6 the yields, the signal to background ratios, the significances, the fit results for the background fraction, the mass and the width of the B_s are listed for the binned mass fits applied to the network selection and the cut based selection of the data in the wide mass range. A complete compilation of the binned mass fits for all studied decay channels can be reviewed in Appendix D.1.1.

The results obtained by the binned fits clearly indicate the better performance of the neural network selection compared to the cut based selection. The figure of merit, the significance $\frac{N_S}{\sqrt{N_S + N_B}}$, is improved by about 15-90%.

Because all unbinned maximum likelihood fits are performed in the narrow mass range, the results for the background fractions obtained by the binned mass fits are recalculated. The new fractions are listed in the tables 6.8 and 6.9 and serve as fixed input quantities to the unbinned maximum likelihood fits.

	$B_s \rightarrow D_s \pi$ modes		
	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$
cut on nnout	0.68	0.92	0.48
N_S (neural net)	1871 ± 48	969 ± 42	722 ± 40
N_S (cut based)	1505 ± 46	711 ± 34	551 ± 36
N_B (neural net)	211 ± 16	396 ± 27	590 ± 18
N_B (cut based)	281 ± 17	318 ± 22	534 ± 16
N_S/N_B (neural net)	8.9 ± 0.7	2.5 ± 0.2	1.2 ± 0.1
N_S/N_B (cut based)	5.4 ± 0.4	2.2 ± 0.2	1.0 ± 0.1
$N_S/\sqrt{N_S + N_B}$ (neural net)	41.0 ± 0.7	26.2 ± 1.0	19.9 ± 0.9
$N_S/\sqrt{N_S + N_B}$ (cut based)	35.6 ± 0.8	22.2 ± 0.9	16.7 ± 0.9
B_s mass [MeV]	5366.4 ± 0.6	5367.8 ± 0.9	5365.8 ± 1.3
B_s width [MeV]	20.9 ± 0.6	20.3 ± 1.1	20.6 ± 1.3
f_b	0.857 ± 0.003	0.912 ± 0.004	0.949 ± 0.003
fraction Λ_b/B_s	-	9.2%	6.2%
fraction B^0/B_s	-	6.7%	-

Table 6.5: Results of the binned fits for the $B_s \rightarrow D_s \pi$ modes in the wide mass range

	$B_s \rightarrow D_s 3\pi$ modes		
	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$
cut on nnout	0.90	0.99	0.99
N_S (neural net)	616 ± 29	260 ± 21	197 ± 24
N_S (cut based)	427 ± 36	184 ± 23	-
N_B (neural net)	272 ± 16	159 ± 13	160 ± 13
N_B (cut based)	1152 ± 34	335 ± 18	-
N_S/N_B (neural net)	2.3 ± 0.2	1.6 ± 0.2	1.2 ± 0.2
N_S/N_B (cut based)	0.4 ± 0.1	0.6 ± 0.1	-
$N_S/\sqrt{N_S + N_B}$ (neural net)	20.68 ± 0.8	12.7 ± 0.9	10.4 ± 1.1
$N_S/\sqrt{N_S + N_B}$ (cut based)	10.7 ± 0.9	8.1 ± 1.0	-
B_s mass [MeV]	5366.6 ± 0.7	5361.7 ± 1.5	5366.3 ± 1.4
B_s width [MeV]	13.2 ± 0.7	15.9 ± 1.7	12.7 ± 1.6
f_b	0.925 ± 0.004	0.932 ± 0.005	0.971 ± 0.004
fraction Λ_b/B_s	-	9.5%	17.3%
fraction B^0/B_s	-	8.5%	25.2%

Table 6.6: Results of the binned fits to the $B_s \rightarrow D_s 3\pi$ modes in the wide mass range

cuts on	$B_s \rightarrow D_s\pi$ modes			$B_s \rightarrow D_s3\pi$ modes	
	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$
$\chi_{r\phi}^2(B_s)$	< 15	< 8	< 6	< 15	< 8
$\chi_{r\phi}^2(D_s)$	< 14	< 15	< 15	< 15	< 10
$\sigma_{Lxy}(B_s)$ [cm]	< 0.04	< 0.04	< 0.04	< 0.04	< 0.04
$\frac{L_{xy}}{\sigma_{Lxy}}(B_s)$	> 7	> 9	> 13	> 11	> 18
$\frac{L_{xy}}{\sigma_{Lxy}}(D_s)$	> 2	> 2	> 2	> 11	> 15
$L_{xy}(B_s \leftarrow D_s)$ [cm]	> -0.02	> -0.02	> -0.01	> -0.005	> 0.01
$ d_0(B_s) $ [cm]	< 0.006	< 0.006	< 0.006	< 0.007	< 0.005
$p_t(B_s)$ [GeV]	> 5.5	> 5.5	> 6.0	> 6.0	> 5.0
$p_t(\pi_{B_s})$ [GeV]	> 1.2	> 1.3	> 1.5	-	-
$\min p_t$ [GeV]	> 0.35	> 0.40	> 0.40	> 0.45	> 0.35
m_{KK} [GeV]	[1.013,1.028]	-	-	[1.010,1.031]	-
$ \Delta m(K\pi, K^*) $ [GeV]	-	< 0.055	-	-	< 0.050
$ \Delta m(KK\pi, D_s) $ [GeV]	-	< 0.025	-	-	-
$ \Delta m(K\pi\pi, D^-) $ [GeV]	-	> 0.024	-	-	> 0.024
$ \Delta m(K\pi\pi, K\pi) $ [GeV]	-	-	> 0.160	-	-
$\Delta R(D_s, \pi_{B_s})$	-	< 1.5	< 1.5	-	-

Table 6.7: Overview of the cuts used in the cut based analysis for all $B_s \rightarrow D_s(3)\pi$ decay modes. For the decay $B_s \rightarrow D_s3\pi$, $D_s \rightarrow 3\pi$ no cut based analysis exists.

	$B_s \rightarrow D_s\pi$ modes		
	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$
f_b	0.332	0.601	0.771
fraction Λ_b/B_s	-	8.8%	5.7%
fraction B^0/B_s	-	5.3%	-

Table 6.8: Recalculated fractions of the $B_s \rightarrow D_s\pi$ decay modes in the narrow mass range.

	$B_s \rightarrow D_s3\pi$ modes		
	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$
f_b	0.599	0.717	0.699
fraction Λ_b/B_s	-	7.9%	5.9%
fraction B^0/B_s	-	7.4%	0.4%

Table 6.9: Recalculated fractions of the $B_s \rightarrow D_s3\pi$ decay modes in the narrow mass range.

To validate the proper implementation of the probability density function modeling the mass space in the fitting framework toy experiments are used. In each toy experiment an unbinned maximum likelihood fit to simulated data is performed. Because the same general function is utilized in all the decay modes, the input parameters for the simulation of the toy experiments are taken from the binned fit function of the $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$ decay. Two different sets of simulated data are used for the toy experiments. One consists of 3000 simulated events to reflect the statistics of a $B_s \rightarrow D_s(3)\pi$ decay in the narrow mass range. The other has 30000 simulated events in order to remove possible statistical fluctuations. The toy experiments are repeated 5000 times and afterwards the deviation of the true input parameter to the fitted parameter normalized to the error of the fitted parameter is evaluated. This quantity is called pull. The pull distribution should be consistent with a Gaussian function with mean zero and width one, if the fit is unbiased. The results obtained for a Gaussian fit to the pull distributions are listed in tables 6.10. They are compatible with the expectation and one can conclude that the fit results are unbiased. In figures 6.13 and 6.14 examples of pull distributions are displayed.

		Number of events per toy experiment	
		3000	30000
B_s mass	Mean	-0.008 ± 0.014	0.035 ± 0.014
	Width	0.995 ± 0.010	0.982 ± 0.010
B_s width	Mean	-0.083 ± 0.014	-0.065 ± 0.014
	Width	0.948 ± 0.010	0.971 ± 0.010

Table 6.10: Results of the toy Monte Carlo experiments for the pull distributions of the fitted mass and width of the B_s meson

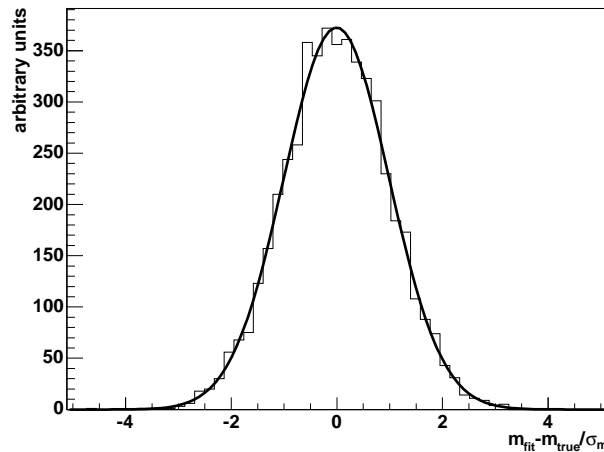


Figure 6.13: Pull distribution of the mass and Gaussian fit for 5000 toy experiments with 3000 events per experiment.

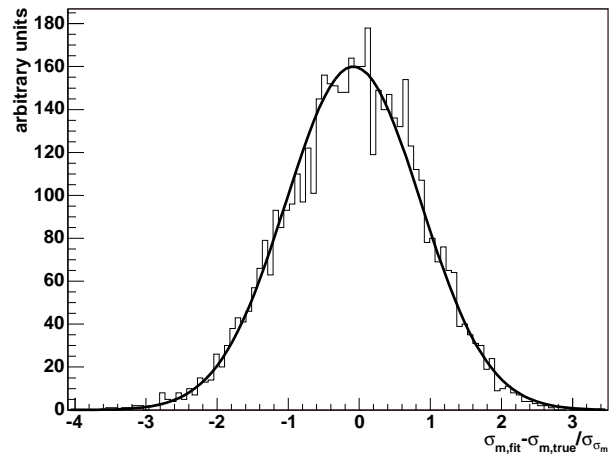


Figure 6.14: Pull distribution of the width and Gaussian fit for 5000 toy experiments with 3000 events per experiment.

Chapter 7

Lifetime Fit

The determination of the proper decay time and its resolution for the different B_s decay modes is a necessary requirement for the mixing fit, because the mixing corresponds to the modulation of the proper decay time distribution. This chapter summarizes the properties of the proper decay time and its resolution. It describes the unbinned maximum likelihood fit simultaneously performed in the lifetime and mass space. Finally the fit results of the studied decay channels of the B_s meson are presented.

7.1 Proper Decay Time

The proper decay time of the B_s is defined as the decay time in its rest frame. It is obtained by measuring the decay length in the x - y plane L_{xy} and correcting this length for the B_s Lorentz boost in the laboratory frame by $\frac{1}{\beta\gamma} = \frac{m_{B_s}}{p_t}$. In this analysis the world average value for the B_s mass is taken, $m_{B_s}^{PDG} = 5.369$ GeV [16]. The formula for the proper decay time corresponds to

$$t = \frac{L_{xy} m_{B_s}^{PDG}}{c p_t} \quad (7.1)$$

The resolution of the proper decay time can be written as

$$\sigma_t = \left(\frac{m_{B_s}^{PDG}}{c p_t} \right) \sigma_{L_{xy}} \oplus \left(\frac{c t}{p_t} \right) \sigma_{p_t} \quad (7.2)$$

The first term describes the quality of the decay length measurement, which is mainly determined by the vertex resolution and the decay topology. This analysis uses an event-by-event reconstruction of the primary vertex instead of using an average beam position. Thereby the decay length resolution is improved. The second term depends on the momentum resolution of each candidate. The contribution of the momentum resolution is small compared to the decay length resolution when studying fully reconstructed hadronic decays modes. This argument does not hold for semileptonic decays, because they have a bad momentum resolution due to the missing neutrino momentum. In this thesis the momentum resolution term is neglected and the expression for the decay time resolution reduces to

$$\sigma_t = \frac{m_{B_s}^{PDG}}{c p_t} \sigma_{L_{xy}} \quad (7.3)$$

7.1.1 Scale Factor Determination for the Proper Decay Time Resolution

The problem of the determination of the scale factor for the proper decay time resolution arises due to the fact that not all properties concerning the vertex reconstruction are understood. The proper decay time resolution is calculated from the uncertainty on the B_s vertex and the event-by-event primary vertex of the $p\bar{p}$ interaction. From samples unbiased in decay length, e.g. data collected with the J/ψ di-muon trigger, an underestimation of the proper decay length resolution is observed. Usually the correction of this underestimation is included in the fitting procedures by determining a scale factor from sidebands that contain a large fraction of prompt events. In the case of the Two Track Trigger these contributions are largely suppressed. Therefore the scale factor is obtained from prompt $D(3)\pi$ samples that share most of the properties of the Two Track Trigger data samples. To extract the scale factor a sequence of several procedures is applied to the data. The calibration sample is binned in different variables on which the scale factor has large dependencies. These variables are $\Delta R(D, \pi)$, the isolation I , η , z , χ^2 for the $D\pi$ topology and I , z , χ^2 for the $D3\pi$ topology. $\Delta R(D, \pi)$ is the distance between the D track and the π track in the ϕ - η plane and can be calculated via the relation

$$\Delta R(D, \pi) = \sqrt{\Delta\phi(D, \pi)^2 + \Delta\eta(D, \pi)^2} \quad (7.4)$$

The isolation is defined as the ratio of the transverse momentum of the B_s to the sum of all transverse momenta of tracks inside a cone with $\Delta R(B_s, \text{track})$

$$I = \frac{p_t(B_s)}{\sum_i p_t(\text{track}_i)} \quad (7.5)$$

In this thesis $\Delta R(B_s, \text{track})$ is 0.7 and the tracks inside the cone are required to have a minimum p_t of 0.35 GeV. The scale factor variation with respect to one variable is fitted by a parabola and the parabolic correction is applied to the proper decay time resolution. This is repeated for each variable until the scale factor is consistent with 1.0. The correction procedure follows the ordering of the variables as listed above and the event-by-event scale factors can be expressed via the following formulas.

For the $D\pi$ topology

$$\begin{aligned} s = & (p_0^{\Delta R} + p_1^{\Delta R} \Delta R + p_2^{\Delta R} \Delta R^2) \\ & \cdot (p_0^I + p_1^I I + p_2^I I^2) \\ & \cdot (p_0^\eta + p_1^\eta \eta + p_2^\eta \eta^2) \\ & \cdot (p_0^z + p_1^z z + p_2^z z^2) \\ & \cdot (p_0^{\chi^2} + p_1^{\chi^2} \chi^2 + p_2^{\chi^2} \chi^2) \end{aligned} \quad (7.6)$$

For the $D3\pi$ topology

$$\begin{aligned} s = & (p_0^I + p_1^I I + p_2^I I^2) \\ & \cdot (p_0^z + p_1^z z + p_2^z z^2) \\ & \cdot (p_0^{\chi^2} + p_1^{\chi^2} \chi^2 + p_2^{\chi^2} \chi^2) \end{aligned} \quad (7.7)$$

The parameters p for the different datasets used in this analysis are listed in the tables 7.1 to 7.4 . Examples of the scale factor distributions for different datasets are displayed in the figures 7.1 and 7.2. The need for two sets of parameters for the scale factor determination has its origin in the two different versions of the reconstruction software used in the `xbhd0d` and `xbhd0h+i` datasets. The average scale factor for signal like events is about 1.4-1.6 for the hadronic B_s decay modes, which is consistent with the scale factor one can determine in the J/ψ data. Therefore the uncorrected proper decay time resolution in the Two Track Trigger data is underestimated. The scale factor obtained by the procedure described above is applied to the proper decay time resolution at all times. The new proper decay time resolution is

$$\sigma'_t = s \cdot \sigma_t \quad (7.8)$$

In the following the prime is dropped and the scaled proper decay time resolution is denoted with σ_t .

	ΔR	I	η	z	χ^2
p_0	1.853	1.259	0.9790	0.99469	0.91398
p_1	-0.642	-0.519	0.0046	-0.00062	0.02723
p_2	0.217	-0.200	0.0928	0.00001	-0.00055

Table 7.1: Parameters p for the scale factor of the $D_s\pi$ decay modes, `xbhd0d` dataset [81].

	ΔR	I	η	z	χ^2
p_0	1.275	1.369	0.988	0.995	1.262
p_1	-0.281	-0.802	0.012	-0.001	0.038
p_2	0.053	0.368	0.063	0.00001	-0.001

Table 7.2: Parameters p for the scale factor of the $D_s\pi$ decay modes, `xbhd0h` and `xbhd0i` datasets [82].

	I	z	χ^2
p_0	1.442	0.99007	0.85712
p_1	0.355	-0.00016	0.02806
p_2	-0.409	0.00002	-0.00071

Table 7.3: Parameters p for the scale factor of the $D_s3\pi$ decay modes, `xbhd0d` dataset [81].

	I	z	χ^2
p_0	1.379	0.995	1.219
p_1	-0.910	-0.001	0.035
p_2	0.487	0.00001	-0.001

Table 7.4: Parameters p for the scale factor of the $D_s3\pi$ decay modes, `xbhd0h` and `xbhd0i` datasets [82].

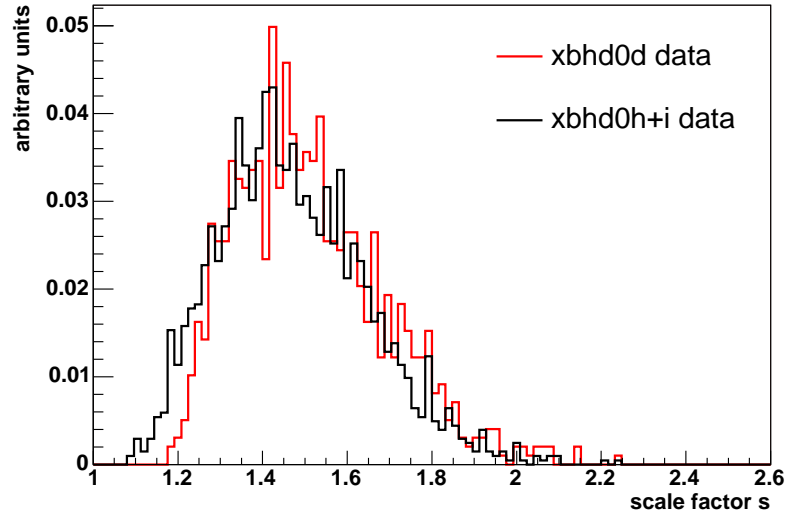


Figure 7.1: Scale factor distributions for signal like events of the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$.

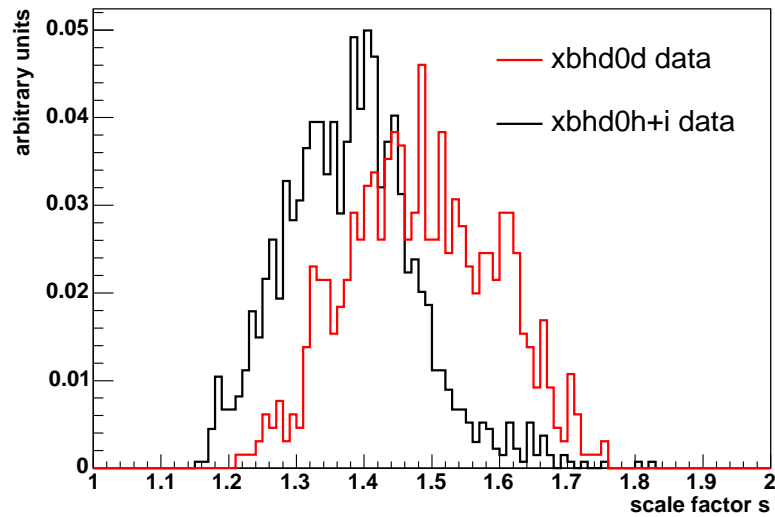


Figure 7.2: Scale factor distributions for signal like events of the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$.

7.2 The Lifetime Fit

The lifetime probability function can be written as

$$P(t) = (1 - f_b)P_S(t) + f_b P_B(t) \quad (7.9)$$

with f_b being the background fraction. It is designed in a general way like the mass probability density function for the use in all $B_s \rightarrow D_s(3)\pi$ decay modes.

The probability density function for the true lifetime is described by an exponential decay

$$P_S^{True}(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \quad (7.10)$$

To include the finite resolution of the detector one has to convolute the true life time distribution with the detector resolution approximated by a Gaussian function centered around zero and width σ_t

$$P_S(t) = e^{-\frac{t}{\tau}} \otimes \mathcal{G}(t - t', \sigma_t) \quad (7.11)$$

Because of the cuts on the impact parameter and decay length during the Two Track Trigger selection, a bias in the lifetime measurement is introduced. The bias is further enhanced due to cuts on variables depending on d_0 and L_{xy} during the preselection of data and later by the neural network selection. To correct for this bias one has to determine an efficiency function as a function of the proper decay time. The efficiency function can be derived from a Monte Carlo sample. It is the ratio of the proper decay time distribution after trigger decision and cuts are applied and the sum over the expected proper decay time distribution for all generated events.

The expected proper decay time distribution is described by the convolution of the true proper decay time with the event-by-event resolution. The efficiency function can be written as

$$\epsilon(t) = \frac{g(t) \text{ after trigger and cuts}}{\sum_i e^{-\frac{t}{\tau}} \otimes \mathcal{G}(t - t', \sigma_{t,i})} \quad (7.12)$$

The complete description of the signal proper lifetime distribution arises to

$$P_S(t) = \left(e^{-\frac{t}{\tau}} \otimes \mathcal{G}(t - t', \sigma_t) \right) \cdot \epsilon(t) \quad (7.13)$$

To determine an analytical expression for the normalization of equation 7.13, the efficiency function is fitted with the following phenomenological description

$$\epsilon(t) = \sum_{i=0}^2 N_{\epsilon,i} \cdot e^{-\frac{t}{\tau_i}} (\beta_i + t)^2 \cdot \theta(t - \beta_i) \quad (7.14)$$

The main drawback of this phenomenological modeling is the difficulty to find a reasonable minimum. In order to obtain fit results the fit is repeated several times with random starting values and the set of parameters with the best χ^2/ndf for the fit function is chosen. In figure 7.3 an example of a fit to the efficiency function is shown for the decay $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$. The efficiency functions of all channels can be found in Appendix C.2.1 .

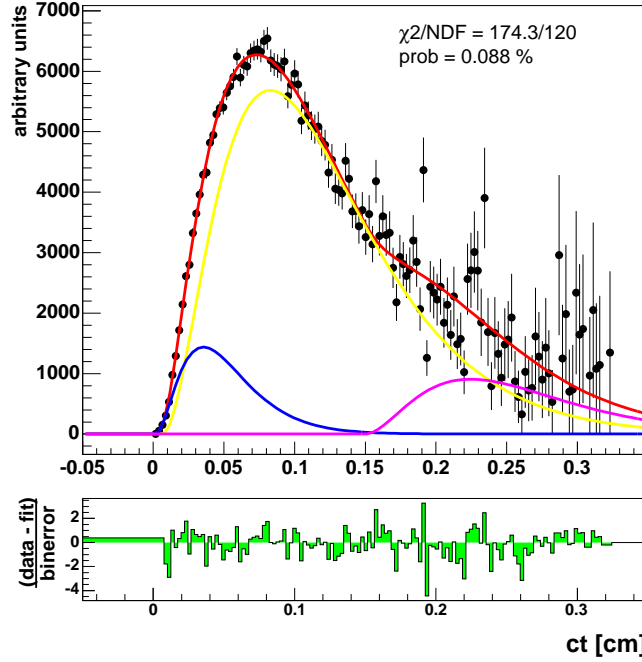


Figure 7.3: Example of the parameterized fit to the efficiency function for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$.

The description of the probability density function for background events in the lifetime space is structured the same way as in the mass space.

$$P_B(t) = f_{comb} \cdot P_{comb} + f_{part} \cdot (1 - f_{comb}) \cdot P_{part} + (1 - f_{part}) \cdot (1 - f_{comb}) \cdot (f_\Lambda P_{\Lambda_b} + (1 - f_\Lambda) \cdot P_{B^0}) \quad (7.15)$$

Because all unbinned fits are performed in the narrow mass range, the contributions of the partially reconstructed decays are neglected. The background probability density function then reduces to

$$P_B(t) = f_{comb} \cdot P_{comb} + (1 - f_{comb}) \cdot (f_\Lambda P_{\Lambda_b} + (1 - f_\Lambda) \cdot P_{B^0}) \quad (7.16)$$

All background contributions are modeled by an exponential decay convoluted with a Gaussian function

$$P_{comb, \Lambda_b, B^0} = N_{comb, \Lambda_b, B^0} \cdot \left(e^{-\frac{t}{\tau_{comb, \Lambda_b, B^0}}} \otimes \mathcal{G}(t, \mu_{comb, \Lambda_b, B^0}, \sigma_{t_{comb, \Lambda_b, B^0}}) \right) \quad (7.17)$$

with N_{comb, Λ_b, B^0} being the normalization.

This description is purely phenomenological and does not reflect any physical properties, e.g. $\tau_{comb, \Lambda_b, B^0}$ is not the proper lifetime of the particle. Figures 7.4 to 7.6 show fits to the different background contributions of the lifetime obtained from Monte Carlo samples (Λ_b , B^0) and the upper sideband (combinatorial background) for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$. The fit results for the other decay modes are given in Appendix C.2.2.

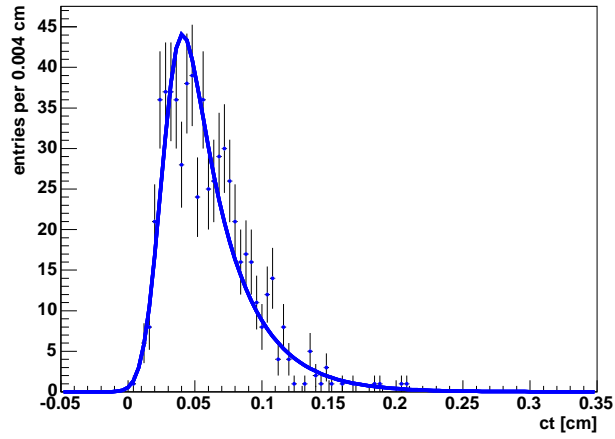


Figure 7.4: Fit of P_{Λ_b} to the Λ_b MC template for the decay $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$

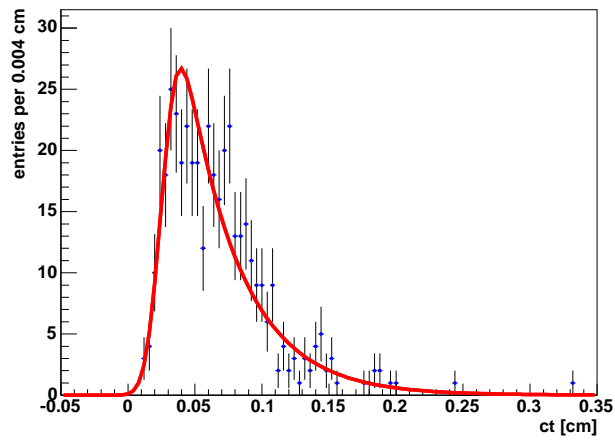


Figure 7.5: Fit of P_{B^0} to the B^0 MC template for the decay $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$

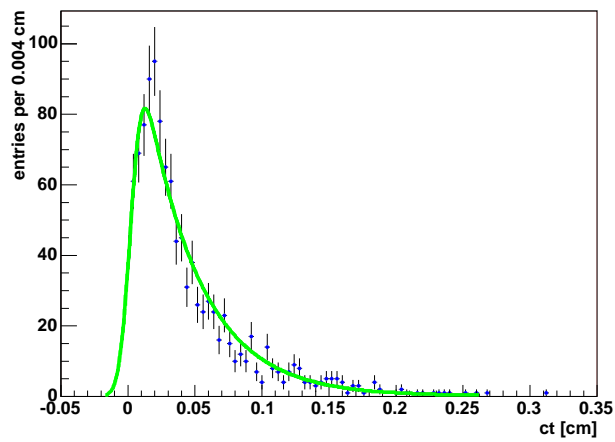


Figure 7.6: Fit of P_{comb} to the upper sideband for the decay $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$

To prepare the input parameters concerning the lifetime of the mixing fit, one performs a simultaneous fit in lifetime and mass space. Including the probability density function of the mass in the unbinned lifetime fits one receives additional information that allows for a better determination of the lifetime. Because the mass and lifetime variables are regarded as uncorrelated, the probability density functions of the signal and background contributions for mass and lifetime space decouple. The maximum likelihood function can be expressed by

$$L = \prod_i P(m_i, t_i) = \prod_i (1 - f_b) \cdot P_S(t_i) \cdot P_S(m_i) + f_b \cdot P_B(t_i) \cdot P_B(m_i) \quad (7.18)$$

The likelihood functions fulfill the normalization relation for every event

$$\int_m \int_t P(m, t) dm dt = 1 \quad (7.19)$$

The parameters for the Λ_b and the B^0 lifetime contributions obtained by the fits to the background Monte Carlo samples are kept fixed for the unbinned fits. Because the parameterization of the lifetime distribution of the upper side band background is not optimal, see figure 7.6, the parameters of the fit to the upper sideband serve as start values for the unbinned fits. All other parameters and fractions are derived of the mass fits and are fixed in the simultaneous unbinned maximum likelihood fit in mass and lifetime space. The fit results are listed in tables 7.5 and 7.6. The measured lifetimes are compatible with the world average values for the flavor specific lifetimes determined by the Heavy Flavor Averaging Group (HFAG) [83]. In figures 7.7 and 7.8 projections of the fit results to the lifetime and mass space are displayed for the decay $B_s \rightarrow D_s \pi, D_s \rightarrow K^* K$. A complete compilation of the fit results for all decay modes can be found in Appendix D.2.1.

To probe the implementation of the unbinned maximum likelihood fit of the lifetime in the fitter framework, a toy experiment study has been performed. The results for the toy experiment studies with 5000 toy experiments and each experiment containing 3000, respectively 30000 events, are summarized in table 7.7. The pull distributions are in good agreement with the expectation of a normal distribution with mean zero and width one.

	$B_s \rightarrow D_s \pi$ modes		
	$D_s \rightarrow \phi \pi$	$D_s \rightarrow K^* K$	$D_s \rightarrow 3\pi$
$B_s ct$ [μm]	449.7 ± 13.8	467.7 ± 22.0	402.3 ± 24.2
B_s mass [MeV]	5366.2 ± 0.6	5367.7 ± 1.0	5365.0 ± 1.3
B_s width [MeV]	21.1 ± 0.5	21.5 ± 0.9	22.1 ± 1.3

Table 7.5: Results of the unbinned lifetime fits for the $B_s \rightarrow D_s \pi$ modes.

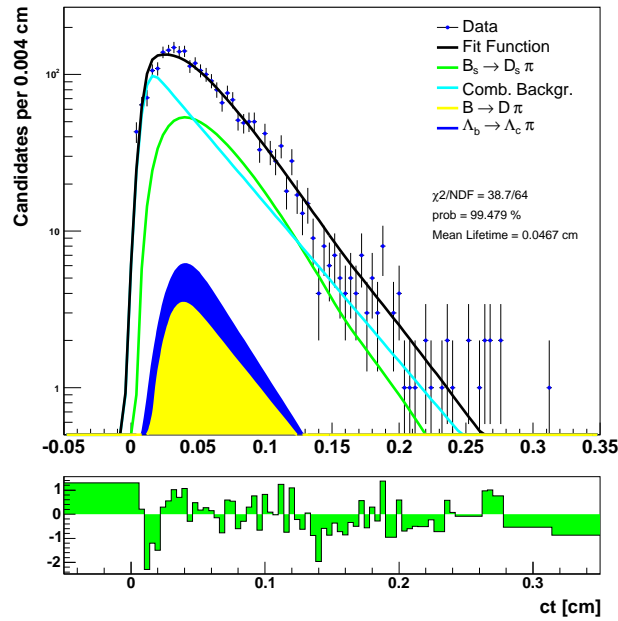


Figure 7.7: Projection of the unbinned fit result to the lifetime space for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

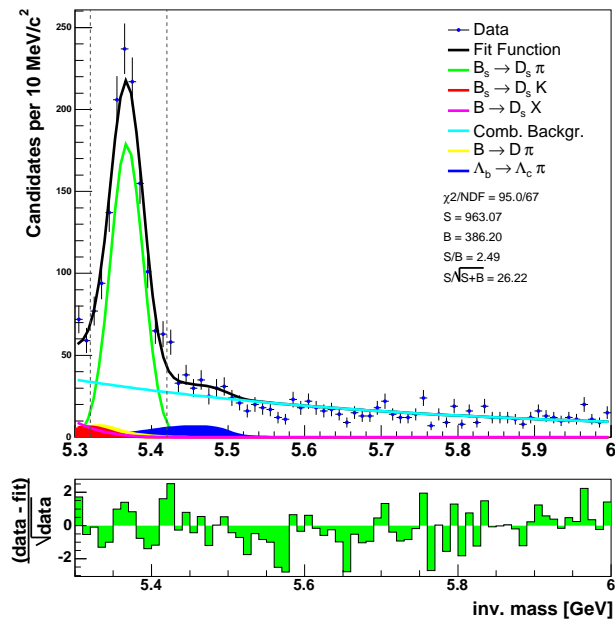


Figure 7.8: Projection of the unbinned fit result to the mass space for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

	$B_s \rightarrow D_s 3\pi$ modes		
	$D_s \rightarrow \phi\pi$	$D_s \rightarrow K^*K$	$D_s \rightarrow 3\pi$
B_s ct [μm]	461.7 ± 23.1	481.4 ± 38.5	387.3 ± 41.5
B_s mass [MeV]	5366.4 ± 0.7	5362.4 ± 1.4	5366.3 ± 1.3
B_s width [MeV]	15.5 ± 0.7	14.9 ± 1.4	12.9 ± 1.2

Table 7.6: Results of the unbinned lifetime fits for the $B_s \rightarrow D_s 3\pi$ modes.

		Number of events per toy experiment	
		3000	30000
B_s ct	Mean	-0.028 ± 0.014	-0.069 ± 0.014
	Width	1.002 ± 0.010	0.993 ± 0.010
B_s mass	Mean	0.017 ± 0.014	0.013 ± 0.014
	Width	1.009 ± 0.010	0.980 ± 0.010
B_s width	Mean	-0.027 ± 0.014	-0.076 ± 0.014
	Width	0.960 ± 0.010	0.960 ± 0.010

Table 7.7: Results of the toy Monte Carlo experiments for the pull distributions of the proper decay time, mass and width of the B_s meson.

Chapter 8

Mixing Fit

In the previous chapters all necessary ingredients for the B_s mixing analysis have been described. These components are now put together to form the probability density function used in the mixing likelihood function and in the amplitude scan method.

8.1 The Mixing Likelihood

The probability density function as mentioned in section 4.1 consists of the mixing terms including the lifetime description and the mass terms.

$$P = (1 - f_b) \cdot P_S(\xi, t, \sigma_t) \cdot P_S(m) + f_b \cdot P_B(\xi, t) \cdot P_B(m) \quad (8.1)$$

The mass term $P(m)$ for signal and background contributions has been subject to chapter 6 and remains unchanged in the mixing likelihood. But the lifetime term $P(t, \sigma_t)$ used in chapter 7 for the probability density of the lifetime space has to be modified to include the oscillation of the B_s meson in the exponential decay. The mixing term for the oscillating signal is expressed by

$$P_S(\xi, t, \sigma_t) = \frac{1}{N_S} \left(\frac{1 + \xi D \cos(\Delta m_s t)}{1 + |\xi|} \cdot e^{-\frac{t}{\tau}} \right) \otimes \mathcal{G}(t - t', \sigma_t) \cdot \epsilon(t) \quad (8.2)$$

with ξ denoting if the B_s meson has mixed or not. This is calculated by multiplying the tagging decision by the charge of the pion (three pion vertex). D is the predicted tagging dilution, N_S the normalization and Δm_s corresponds to the mixing frequency of the B_s meson. The analytic expression for the normalization N_S can be found in [84].

Assuming the background components in the upper sideband do not oscillate, the descriptions of tagging and lifetime decouple. The predicted dilutions have no meaning for non-oscillating background contributions that do not contain any b quark, e.g. the combinatorial background. To incorporate possible flavor tagging asymmetries for the background a dilution-like fit parameter is introduced, the so called background dilution D_{bg} . The background probability for the tagging and lifetime part can be written under the assumption of an equal number of positively and negatively tagged candidates in the background as

$$P_B(\xi, t) = (1 + \xi D_{bg}) \cdot P_B(t) \quad (8.3)$$

The complete maximum likelihood function for the determination of the mixing frequency Δm_s corresponds to

$$L = \prod_i (1 - f_b) \cdot \frac{1}{N_S} \left(\frac{1 + \xi_i D_i \cos(\Delta m_s t_i)}{1 + |\xi_i|} \cdot e^{-\frac{t_i}{\tau}} \right) \otimes \mathcal{G}(t_i - t', \sigma_{t,i}) \cdot \epsilon(t_i) \cdot P_S(m_i) + f_b \cdot (1 + \xi_i D_{bg}) \cdot P_B(t_i) \cdot P_B(m_i) \quad (8.4)$$

The probability density functions P_S and P_B are normalized to the same value according to

$$\sum_{\xi=-1,0,1} \int_m \int_t P_S(m, t, \xi, \sigma_t) \, dm \, dt = \sum_{\xi=-1,0,1} \int_m \int_t P_B(m, t, \xi) \, dm \, dt \quad (8.5)$$

For the final mixing fit all parameters for the mass and lifetime components determined in the previous fits are kept fixed. The mixing frequency and the background dilution are the only free parameters in the mixing fit. The tagging decision and the predicted dilution of each candidate is given by the same side tagger or the Karlsruhe neural network tagger. The mixing fits are performed separately for each tagging algorithm. One expects the mixing frequency to be 17.77 ps^{-1} [4]. The fits for the mixing frequency are carried out in two different ways. The decay modes are fitted separately as well as in various combinations with each other. For the combined fits the mixing frequency and the background dilution are common parameters to all decay modes included in the fit. The results for the mixing frequency of the combined fit and the fit for the $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ decay mode agree in both applied tagging algorithms. However in the case of the $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$ decay mode the results for the different taggers are not compatible within the uncertainties. The results for this decay mode are also less significant as expected from the large significance in the exclusive reconstruction of the B_s mesons. Therefore the maximum likelihood functions for this channel are evaluated in more detail. It was discovered that a few events have a very large contribution to the log likelihood function and therefore mainly determine the minimum of the log likelihood function. In figure 8.1 the likelihood ratio for such an event with contradicting minima for both taggers is displayed. The likelihood ratio is defined as

$$LR = -2 \ln L(1) - 2 \ln L(0) \quad (8.6)$$

where $L(1)$ denotes the mixing likelihood and $L(0)$ is the null hypothesis, which corresponds to the maximum likelihood function without the mixing term.

The fact that only a few events determine the mixing frequency leads to the conclusion that the maximum likelihood fit is susceptible to statistical fluctuations and wrongly measured or predicted input variables of the likelihood function.

Furthermore an asymmetry concerning the flavor of the B_s meson at decay time has been observed for the $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$ decay mode. In figures 8.2 and 8.3 the difference in the likelihood ratios are visible in the separate B_s and \bar{B}_s samples. The asymmetry is too large to be explained by differences of particles and antiparticles during the reconstruction or in the tagging algorithms and it has not been observed in the other decay modes. Because the neural network selection used for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$, is taken from [5], this decay channel was not studied in detail in this thesis. A further investigation of the $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$ decay channel is under way.

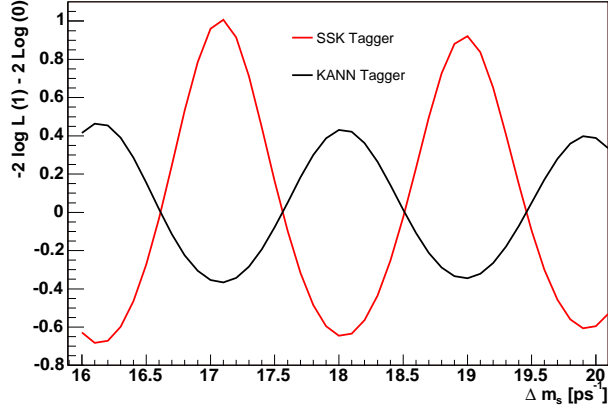


Figure 8.1: Likelihood ratio as a function of Δm_s for a single event (run=199042, event=8793379) for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$.

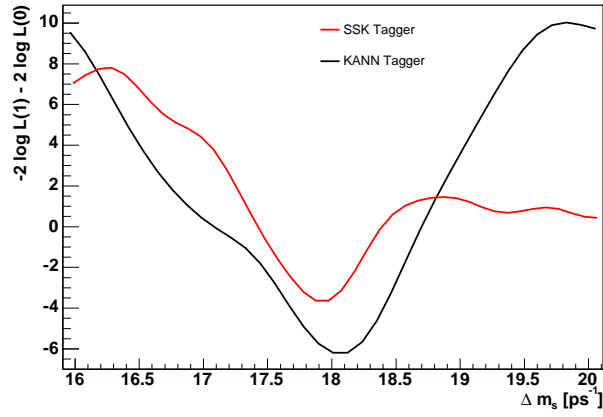


Figure 8.2: Likelihood ratio as a function of Δm_s for the $B_s \rightarrow D_s^- \pi^+$, $D_s^- \rightarrow \phi \pi^-$ decay mode. The charge conjugated decay mode is excluded.

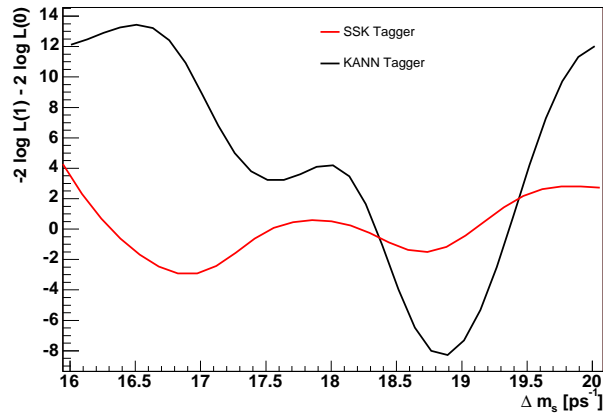


Figure 8.3: Likelihood ratio as a function of Δm_s for the $\bar{B}_s \rightarrow D_s^+ \pi^-$, $D_s^+ \rightarrow \phi \pi^+$ decay mode. The charge conjugated decay mode is excluded.

All results of the unbinned maximum likelihood fit of the mixing frequency are summarized in table 8.1 and graphically represented in figure 8.4. The likelihood ratios of the separate mixing fits for the $B_s \rightarrow D_s\pi$, $D_s \rightarrow \phi\pi$ and $D_s \rightarrow K^*K$ decay modes are displayed in figures 8.5 and 8.6. The likelihood ratios of the combined fits are shown in figure 8.7 and 8.8.

$B_s \rightarrow D_s\pi$ $D_s \rightarrow$	Δm_s [ps^{-1}] SSKT	$D_{bg} \times 10^{-3}$ SSKT	Δm_s [ps^{-1}] KANNT	$D_{bg} \times 10^{-3}$ KANNT
$\phi\pi$	$17.94^{+0.24}_{-0.22}$	-2.8 ± 42.3	$18.69^{+0.17}_{-0.20}$	13.3 ± 36.8
$\phi\pi (B_s)$	$17.92^{+0.20}_{-0.22}$	-18.8 ± 57.9	$18.06^{+0.23}_{-0.24}$	-8.4 ± 50.4
$\phi\pi (\bar{B}_s)$	$16.90^{+0.30}_{-0.32}$	15.6 ± 61.8	18.86 ± 0.16	40.8 ± 53.5
K^*K	17.84 ± 0.17	27.1 ± 33.4	$17.85^{+0.15}_{-0.14}$	43.5 ± 28.7
combined $\phi\pi, K^*K, 3\pi$	$17.89^{+0.13}_{-0.14}$	15.3 ± 17.8	17.86 ± 0.12	12.1 ± 15.5
combined $K^*K, 3\pi$	$17.86^{+0.16}_{-0.18}$	19.2 ± 19.6	17.84 ± 0.12	11.8 ± 17.1

Table 8.1: Results of the mixing fit.

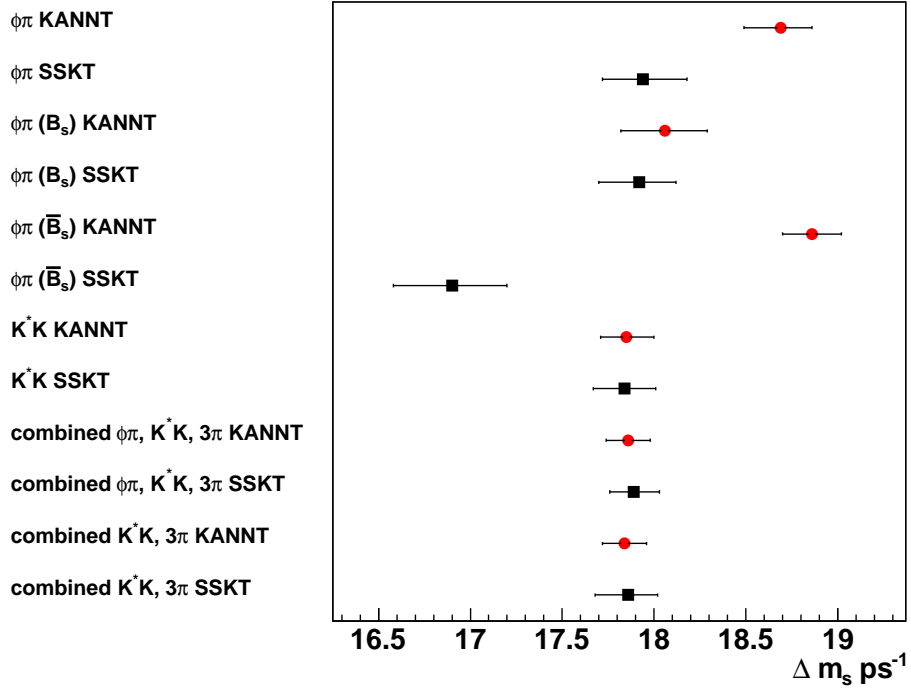


Figure 8.4: Graphical representation of the results obtained by the mixing fit.

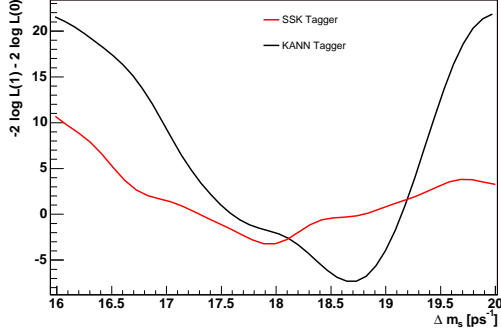


Figure 8.5: Likelihood ratio as a function of Δm_s for the $B_s \rightarrow D_s\pi$, $D_s \rightarrow \phi\pi$ decay mode.

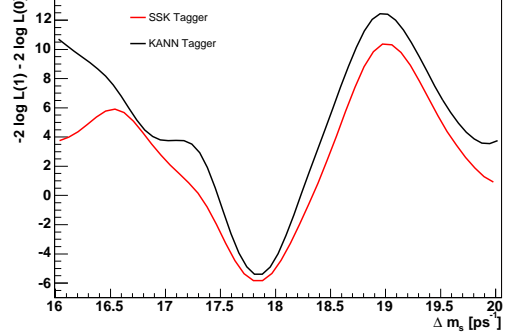


Figure 8.6: Likelihood ratio as a function of Δm_s for the $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K$ decay mode.

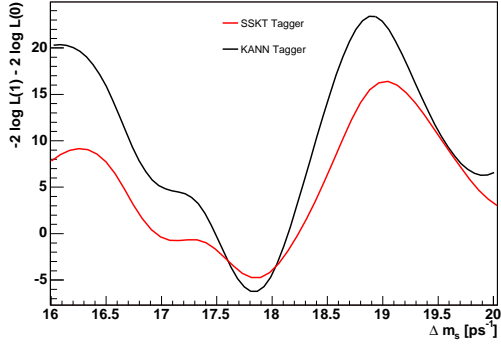


Figure 8.7: Likelihood ratio as a function of Δm_s for the combination of the $B_s \rightarrow D_s\pi$, $D_s \rightarrow K^*K, 3\pi$ decay modes

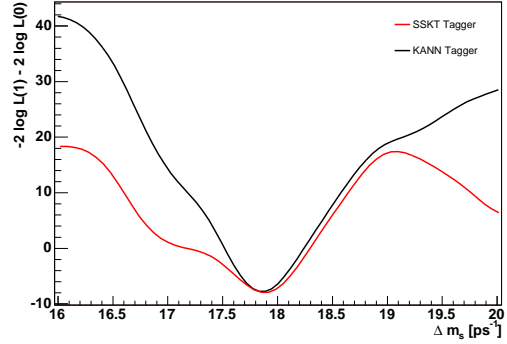


Figure 8.8: Likelihood ratio as a function of Δm_s for the combination of the $B_s \rightarrow D_s\pi$, $D_s \rightarrow \phi\pi, K^*K, 3\pi$ decay modes

The implementation of the fit model in the fitter framework is validated using toy Monte Carlo experiments. The results of the toy Monte Carlo experiments can be found in table 8.2 and the pull distributions are displayed in figure 8.9 and 8.10. The uncertainty on Δm_s in case of toy experiments with 3000 generated events is underestimated, because the negative log likelihood function is asymmetric in the minimum. However the error used to calculate the pull is a MIGRAD error, which assumes a parabolic behavior of the negative log likelihood function in the minimum. The underestimation of the uncertainty vanishes for toy experiments using 30000 generated events as the negative log likelihood function approaches a parabolic shape in its minimum.

		Number of events per toy experiment	
		3000	30000
Δm_s	Mean	-0.008 ± 0.016	-0.018 ± 0.015
	Width	1.104 ± 0.014	1.017 ± 0.011

Table 8.2: Results of the toy Monte Carlo experiments for the pull distributions of the B_s mixing frequency.

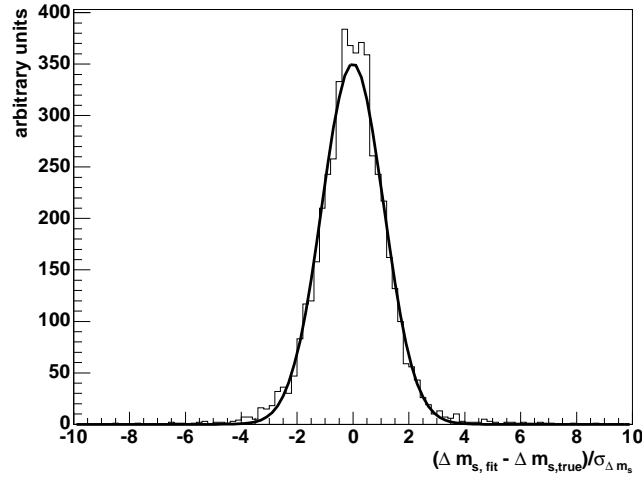


Figure 8.9: Pull distribution of Δm_s and Gaussian fit for 5000 toy experiments with 3000 events per experiment.

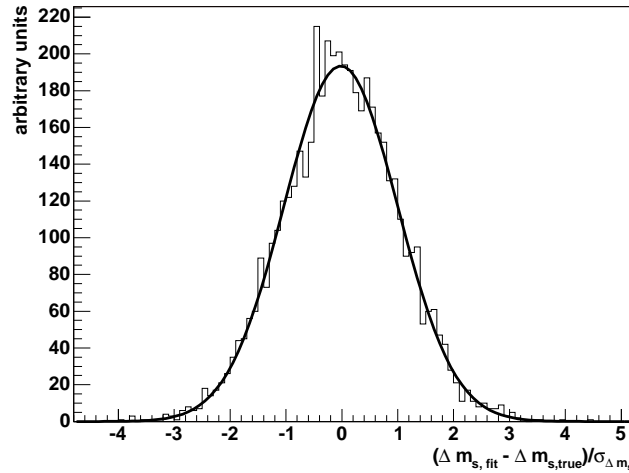


Figure 8.10: Pull distribution of Δm_s and Gaussian fit for 5000 toy experiments with 30000 events per experiment.

8.2 The Amplitude Scan Method

The amplitude scan method provides the opportunity to probe for the mixing frequency in the case of low statistics. Furthermore one can easily combine different measurements and determine the sensitivity and the limits for the results of the mixing analysis. The amplitude scan is based on the idea of scanning over a range of frequencies to determine the mixing frequency. Therefore a Fourier-like coefficient, called amplitude A , is introduced the signal term of the maximum likelihood function

$$P_S(\xi, t, \sigma_t) = \frac{1}{N_S} \left(\frac{1 + \xi A D \cos(\Delta m_s t)}{1 + |\xi|} \cdot e^{-\frac{t}{\tau}} \right) \otimes \mathcal{G}(t - t', \sigma_t) \cdot \epsilon(t) \quad (8.7)$$

During the amplitude scan all parameters are kept fixed in the maximum likelihood function and the amplitude A is determined for the given frequency. The amplitude should be one at the mixing frequency Δm_s and zero otherwise. The shape of the resonance peak in the signal region is expected to follow a Breit-Wigner function. The width of the Breit-Wigner function is determined by the lifetime of the B_s mesons. A bias in the proper decay time distribution introduces an undershoot at both sides of the amplitude peak. The exclusion limit at a 95% confidence level is defined as the largest value of the mixing frequency below which all other frequencies are excluded and it can be expressed by

$$A + 1.645 \cdot \sigma_A < 1 \quad (8.8)$$

The sensitivity at a 95% confidence level for the amplitude scan is defined as

$$1.645 \cdot \sigma_A = 1 \quad (8.9)$$

In figures 8.12 and 8.13 the amplitude scans for the combined fit of the decay modes $B_s \rightarrow \pi, D_s \rightarrow \phi\pi, K^*K, 3\pi$ are displayed for both taggers. The sensitivity of the same side kaon tagger is 25.0 ps^{-1} and 29.2 ps^{-1} for the Karlsruhe neural network tagger. The exclusion limit is 17.0 ps^{-1} for the same side kaon tagger and 17.5 ps^{-1} for the Karlsruhe neural network tagger.

Furthermore the amplitude can be used to calibrate the tagging algorithm. The amplitude is determined in the same way as in the amplitude scan method. But instead of scanning over a range of frequencies, one fixes the mixing frequency to the value obtained in the previous unbinned maximum likelihood fit and only performs the fit once at that frequency. If the predicted dilution of the tagger is over- or underestimated the amplitude will deviate from one at the mixing frequency. It is the first attempt to calibrate tagging algorithms using the B_s oscillation measurement. The calibrated tagging power can be expressed as

$$T = \epsilon D^2 = \epsilon \frac{1}{n} \sum_{i=1}^n (A D_i)^2 \quad (8.10)$$

with n being the number of events in the signal region and D_i the predicted dilution for an event. An example for the distribution of the squared predicted dilution for both tagging algorithms is given in figure 8.11.

The results for the fitted amplitudes and the calibrated tagging power are listed in table 8.3. No systematic uncertainties are given as the statistical uncertainties are dominant in the case of the fitted amplitudes [4]. The problems observed in the results of the unbinned mixing fit of the $B_s \rightarrow D_s\pi$, $D_s \rightarrow \phi\pi$ decay channel for the taggers are propagated to the results of the amplitude scan. The differences in the mixing frequencies for single channels influences the results of a combined amplitude scan. This is due to the fact that the fitted amplitudes at a given frequency add up as follows

$$A = \sum_i \frac{A_i}{\sigma_{A_i}^2} \sigma_A \quad \text{with} \quad \frac{1}{\sigma_A^2} = \sum_i \frac{1}{\sigma_{A_i}^2} \quad (8.11)$$

In the case of the $B_s \rightarrow D_s\pi$, $D_s \rightarrow \phi\pi$ decay mode and the Karlsruhe neural network tagger, the amplitude scan probes at the combined mixing frequency of 17.86 ps^{-1} , yet the frequency determined by the single maximum likelihood fit is 18.69 ps^{-1} . The amplitude for the single fit at this frequency is 0.87 ± 0.31 but 0.57 ± 0.29 at the probe frequency of the combined fit. This explains the low amplitude of the Karlsruhe neural network tagger for the combined fit of all $B_s \rightarrow D_s\pi$ decay modes.

The results obtained for the calibrated tagging power agree with the expectations of about 4% for the same side kaon tagger and about 9% for the neural network tagger. However the large uncertainties prohibit the conclusion which tagging algorithm is superior. The results can be improved by adding more data as up to now the uncertainty of the amplitude is dominated by statistics. In this thesis everything has been prepared to include the missing $B_s \rightarrow D_s 3\pi$ decay modes in the mixing analysis and the tagging calibration. This was prevented by limited availability of computer resources, as the process of implementing the newly developed neural network tagger in the data needs a lot of computing power.

$B_s \rightarrow D_s\pi$ $D_s \rightarrow$	A SSKT	A KANNT	T SSKT	T KANNT
combined $\phi\pi$, K^*K , 3π	0.97 ± 0.33	0.69 ± 0.22	$4.4 \pm 3.0\%$	$4.8 \pm 3.1\%$
combined K^*K , 3π	1.08 ± 0.48	0.84 ± 0.32	$5.7 \pm 5.1\%$	$7.5 \pm 5.7\%$

Table 8.3: The fitted amplitudes and the corresponding tagging powers.

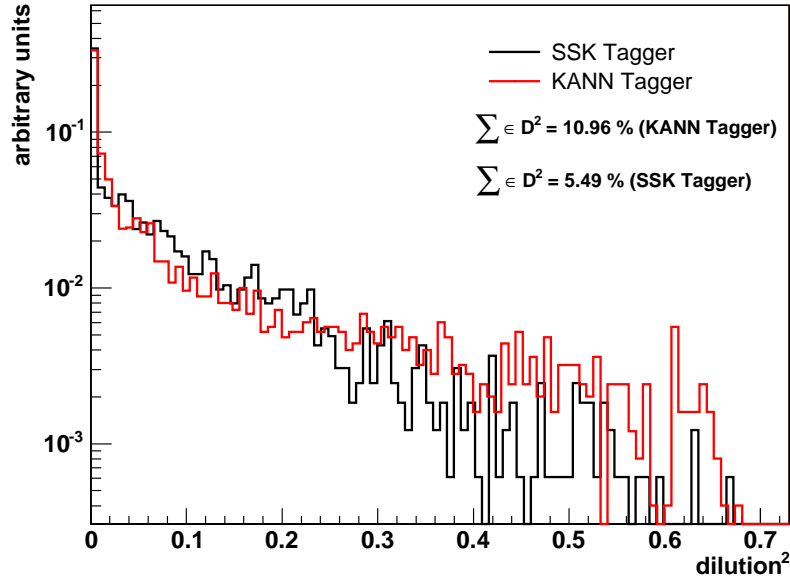


Figure 8.11: Squared predicted dilution of the SSK and KANN tagging algorithms for signal like events of the $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ decay in the narrow mass range.

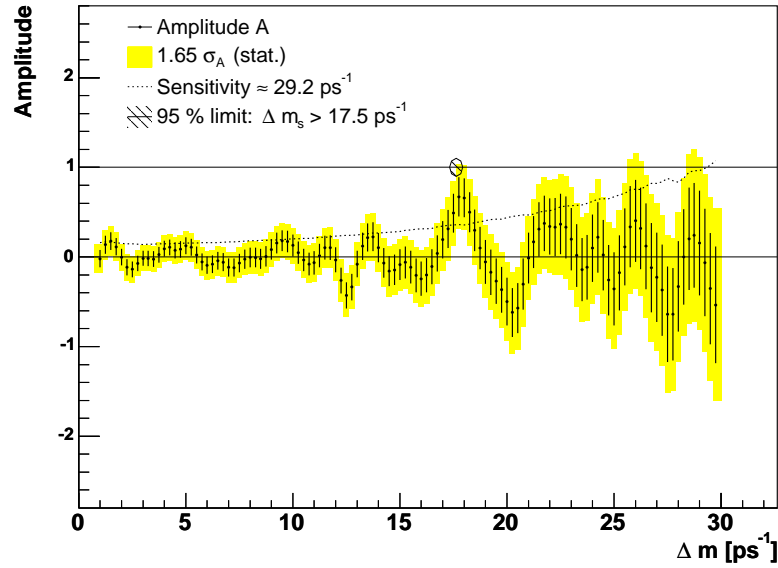


Figure 8.12: Amplitude scan for the combination of $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$, $K^* K$, 3π decay modes using the Karlsruhe neural network tagger.

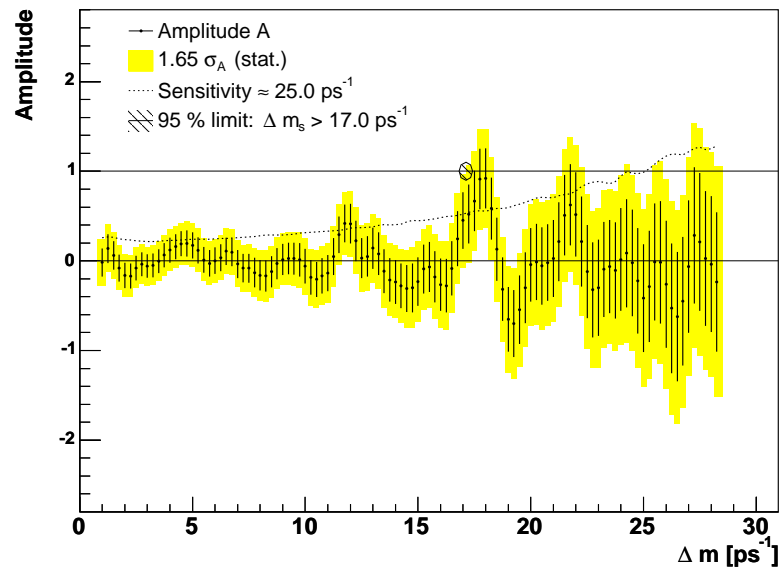


Figure 8.13: Amplitude scan for the combination of $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$, $K^* K$, 3π decay modes using the same side kaon tagger.

Chapter 9

Conclusion and Outlook

The idea of using B_s oscillations to calibrate different tagging algorithms has been the main aim of this thesis. To achieve this goal a complete B_s mixing analysis is necessary. For the full mixing analysis several prerequisites have been met.

- A fitting framework was created with the intention of an easy handling of many decay modes and the corresponding fit functions. The framework is not only used in this thesis, but serves as cornerstone for many other analyses in our group, e.g. the observation of orbitally excited B_s mesons [7].
- The next step towards the measurement of the B_s mixing frequency is the optimization of the selection of exclusive B_s mesons. In the thesis six different fully reconstructed hadronic decay modes of the B_s meson have been studied, namely the decays $B_s \rightarrow D_s(3)\pi$ with $D_s \rightarrow \phi\pi, K^*K, 3\pi$. Thereby previous neural network selections of the decay $B_s \rightarrow D_s\pi$ with $D_s \rightarrow \phi\pi, 3\pi$ and $B_s \rightarrow D_s3\pi$ with $D_s \rightarrow \phi\pi$ taken from [5, 6] have been updated and combined with new selections of the decays $B_s \rightarrow D_s(3)\pi$ with $D_s \rightarrow K^*K$ and $B_s \rightarrow D_s3\pi$ with $D_s \rightarrow 3\pi$. Altogether about 4600 B_s signal events have been found by using neural networks in the final selection compared to about 3400 signal events in the previously used cut based analysis.
- A further requisite for the determination of the B_s mixing frequency is the description of the probability density functions of the mass and lifetime space in the maximum likelihood function of the mixing fit. The values measured for the B_s mass and the B_s lifetime in the studied B_s decay channels are in good agreement with the world average values [16]. In the mixing fit these values are used as fixed input parameters.
- For the necessary b flavor tagging at production time a newly developed Karlsruhe neural network tagging algorithm has been implemented in the analysis and compared to the best available tagging algorithm at the CDF II experiment, the same side kaon tagging algorithm.

The mixing frequency of the B_s meson is determined by the unbinned maximum likelihood mixing fit of the combination of the three decay modes, $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi, K^*K, 3\pi$. The $B_s \rightarrow D_s3\pi$ decay modes are not included in the combined fit as the datasets used in thesis do not yet contain the tagging information of the Karlsruhe neural network tagger.

The final result for the mixing frequency Δm_s of the combined fit is

$$\Delta m_s = 17.89_{-0.14}^{+0.13} \text{ ps}^{-1} \text{ for the same side kaon tagger}$$

$$\Delta m_s = 17.86 \pm 0.12 \text{ ps}^{-1} \text{ for the Karlsruhe neural network tagger}$$

Both results are compatible with the mixing frequency of $17.77 \pm 0.10(\text{stat}) \pm 0.07(\text{sys}) \text{ ps}^{-1}$ measured in [4].

In order to calibrate the different tagging algorithms the amplitude is fitted at the measured mixing frequency. The fitted amplitude determines the tagging power of the flavor tagging algorithm. The tagging power is calculated from the fitted amplitude, the efficiencies and the predicted dilutions of the tagging algorithms. One obtains the following values for the tagging power of the tagging algorithms

$$T = 4.4 \pm 3.0\% \text{ for the same side kaon tagger}$$

$$T = 4.8 \pm 3.1\% \text{ for the Karlsruhe neural network tagger}$$

As the measured mixing frequency in the $B_s \rightarrow D_s$, $D_s \rightarrow \phi\pi$ decay mode is shifted by approximately 0.8 ps^{-1} compared to the mixing frequency determined in the combined fit, the fitted amplitude becomes smaller and the calibrated tagging power is impaired. Therefore the tagging powers are additionally determined in a combined fit excluding the $B_s \rightarrow D_s$, $D_s \rightarrow \phi\pi$ decay mode. The results of the tagging powers improved but have larger uncertainties due to smaller statistics. The results of the tagging powers are

$$T = 5.7 \pm 5.1\% \text{ for the same side kaon tagger}$$

$$T = 7.5 \pm 5.7\% \text{ for the Karlsruhe neural network tagger}$$

This is the first time a calibration of tagging algorithms using B_s oscillations has been attempted and the Karlsruhe neural network tagger has been tested on data. The Karlsruhe neural network tagger is already quite promising yet has to be further evaluated with more data in order to reduce statistical uncertainties.

Furthermore all prerequisites have been met to include the $B_s \rightarrow D_s 3\pi$ decay modes in the calibration of the tagging power as soon as the Karlsruhe neural network tagger is available for these modes. The calibration was done with a data sample of 1 fb^{-1} and the complete mixing analysis including the calibration will be repeated soon on the newly processed data sample of an integrated luminosity of about 2 fb^{-1} . The main effort to improve the tagging power has yet to be the full understanding of the observed problems in the $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi\pi$ decay mode. If the calibrated tagging power of the newly developed Karlsruhe neural network tagger can reach its predicted value of 9.31% the tagging power will be nearly twice as good compared to the tagging power of currently available tagging algorithms. Therefore all analyses using flavor tagged B_s mesons will profit of a doubled statistics. The further improvement of tagging algorithms will help to understand more of the properties of the B_s meson system, e.g. allow for the measurement of mixing induced CP violation. It is the beginning of a new chapter for B_s meson physics which started with the recent measurement of the B_s mixing frequency.

Appendix A

List of Variables

$L_{xy}(P)$ Decay length of the particle P

$L_{xy}/\sigma_{L_{xy}}(P)$ Decay length significance of the particle P

$\sigma_{L_{xy}}(P)$ Error of the decay length of the particle P

$L_{xy}(P \leftarrow C)$ Decay length between Parent P and child C

$\chi_{3D}^2(P)$ Three-dimensional χ^2 of the vertex fit for the particle P

$\chi_{r\phi}^2(P)$ Two-dimensional χ^2 of the vertex fit for the particle P

$|d_0(P)|$ Impact parameter of the particle P

$d_0/\sigma_{d_0}(P)$ Impact parameter significance of the particle P

$\min d_0(\pi_{B_s}^1, \pi_{B_s}^2, \pi_{B_s}^3)$ Minimum of the impact parameters of all pions tracks from the B_s

$\max d_0(\pi_{B_s}^1, \pi_{B_s}^2, \pi_{B_s}^3)$ Maximum of the impact parameters of all pions tracks from the B_s

$\min \frac{d_0}{\sigma_{d_0}}$ Minimum of the impact parameter significances of all tracks

lts. $d_0(P)$ Lifetime signed impact parameter of the particle P (see fig. A.1)

lts. $d_0/\sigma_{d_0}(P)$ Lifetime signed impact parameter significance of the particle P

lts. $d_0/\sigma_{d_0}(\pi_{D_s}^n)$ Lifetime signed impact parameter significance of the n-th (n=1,2,3) pion originating from the D_s meson.

\min lts. d_0/σ_{d_0} Minimum of the lifetime signed impact parameter significances of all tracks

$p_t(P)$ Transverse momentum of the particle P

$\min p_t$ Minimum of the transverse momenta of all tracks

helicity angle K_ϕ^1 Cosine of the angle between the three momentum of the D_s meson $\vec{p}(D_s)$ and the three-momentum $\vec{p}(K^1)$ of the first of the two kaons from the decay $\phi \rightarrow KK$ boosted to the rest frame of the ϕ meson (the ‘‘helicity angle’’)

helicity angle K_{K^*} Cosine of the angle between the three momentum of the D_s meson $\vec{p}(D_s)$ and the three-momentum $\vec{p}(K)$ of the kaon from the decay $K^* \rightarrow K\pi$ boosted to the rest frame of the K^* meson (the ‘‘helicity angle’’)

CMS angle π_{B_s} Angle between the three-momentum of the B_s meson $\vec{p}(B)$ and the three-momentum $\vec{p}(\pi_B)$ of the pion from the decay $D_s \rightarrow \pi\pi\pi$ boosted to the B_s mesons rest frame

$\theta(\vec{n}_{(\pi_{B_s}^1, \pi_{B_s}^2)}, \vec{p}(3\pi))$ Angle between the three-momentum of the three pion vertex and the normal vector of the plane spanned by the momenta of the two pions $\pi_{B_s}^1$ and $\pi_{B_s}^2$, (see fig. A.2)

$$\theta(\vec{n}_{(\pi_{B_s}^1, \pi_{B_s}^2)}, \vec{p}(3\pi_{vtx})) = \angle(\vec{n}, \vec{p}(3\pi_{vtx}))$$

$$\vec{n} = \begin{cases} \vec{p}(\pi_{B_s}^1) \times \vec{p}(\pi_{B_s}^2) & , \text{ if } m_{\pi_{B_s}^1 \pi_{B_s}^2} > m_{\pi_{B_s}^2 \pi_{B_s}^3} \\ -(\vec{p}(\pi_{B_s}^1) \times \vec{p}(\pi_{B_s}^2)) & , \text{ if } m_{\pi_{B_s}^1 \pi_{B_s}^2} \leq m_{\pi_{B_s}^2 \pi_{B_s}^3} \end{cases}$$

$\Delta R(D_s, \pi_{B_s})$ Opening angle (in sterad) between the D_s meson and the π from the decay $B_s \rightarrow D_s\pi$

$q(P)$ Charge of the particle P

inv. mass $m_{P_1 P_2}$ invariant mass of the four momentum $p = p_{P_1} + p_{P_2}$. P^1 and P^2 denote different particles, e.g. $\pi_{D_s}^1$ and $\pi_{D_s}^2$

inv. mass $m_{P_1 P_2}^2$ squared invariant mass of the four momentum $p = p_{P_1} + p_{P_2}$

$\min(m_P^{12}, m_P^{23})$ minimum of the invariant masses of the four momenta $p_{12} = p_{P^1} + p_{P^2}$ and $p_{23} = p_{P^2} + p_{P^3}$

$\min(m_P^{12}, m_P^{23})$ maximum of the invariant masses of the four momenta $p_{12} = p_{P^1} + p_{P^2}$ and $p_{23} = p_{P^2} + p_{P^3}$

inv. mass $m_{3\pi_{B_s}}$ invariant mass of the three pion vertex

m_{diff} ratio of the invariant mass of the three pion vertex and the mass difference of the B_s and the D_s meson $m_{\text{diff}} = \frac{m_{3\pi_{B_s}}}{m_{B_s} - m_{D_s}}$

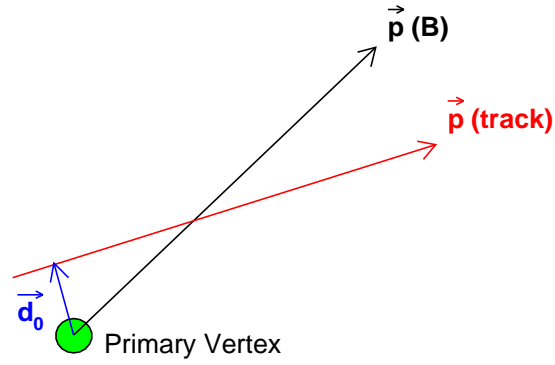


Figure A.1: Definition of the lifetime–signed impact parameters for a track in an event with a B decay : lts. $d_0 = |\vec{d}_0| \text{signum}(\vec{d}_0 \vec{p}_B)$

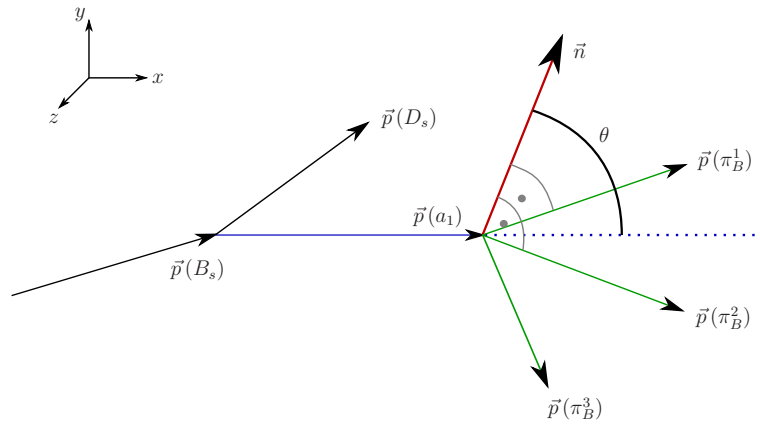


Figure A.2: Definition of the angle $\theta(\vec{n}_{(\pi_{B_s}^1, \pi_{B_s}^2)}, \vec{p}(3\pi))$ for the decay $B_s \rightarrow D_s 3\pi$. Here the 3π vertex is denoted as a_1 .

Appendix B

Compilation of Training Results

This compilation consists of the plots and tables obtained in the training of the B_s channels. They show the correlation matrix, the significance table, the plot of the separation of signal and background and the plot of the purity as a function of network output.

The entries ρ_{ij} of the correlation matrix for all pairs of variables x_i, x_j are calculated using the following quantities as input :

- the covariance matrix C for N variables including the target.

$$C_{ij} = \frac{1}{n} \sum_{events} (x_i - \langle x_i \rangle) \cdot (x_j - \langle x_j \rangle) \quad (\text{B.1})$$

with n the number of training events and $\langle x \rangle$ the expectation value of x

- the variance of the variable x : $V[x_i]$

This leads to the correlation coefficients

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{V[x_i]} \sqrt{V[x_j]}} \quad (\text{B.2})$$

The plot of the purity versus the network output should indicate a linear behavior of the purity for a successful training :

$$P(\text{nnout}) = \frac{(\text{nnout} + 1)}{2} \quad (\text{B.3})$$

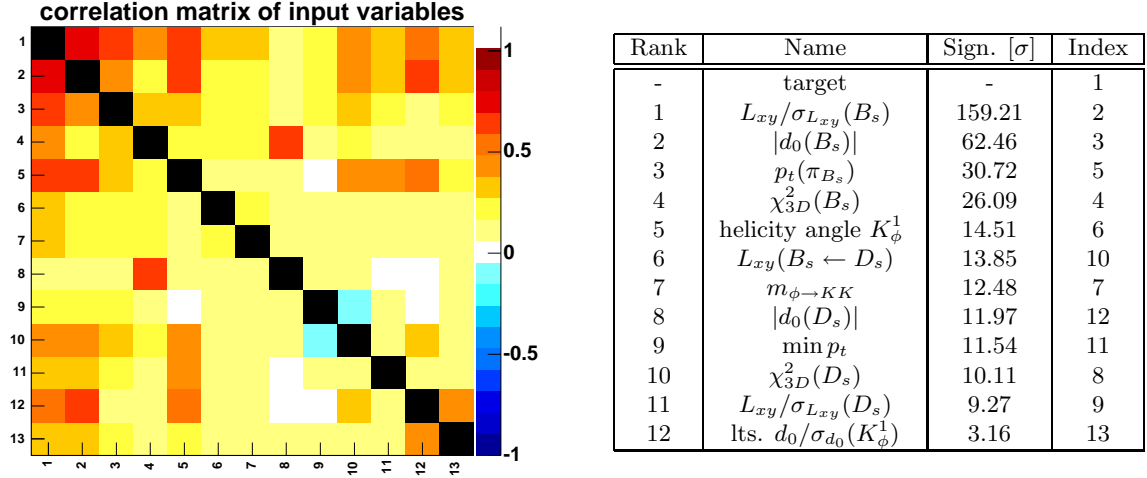


Figure B.1: Correlation matrix and significance table of the training variables for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi, \phi \rightarrow KK$.

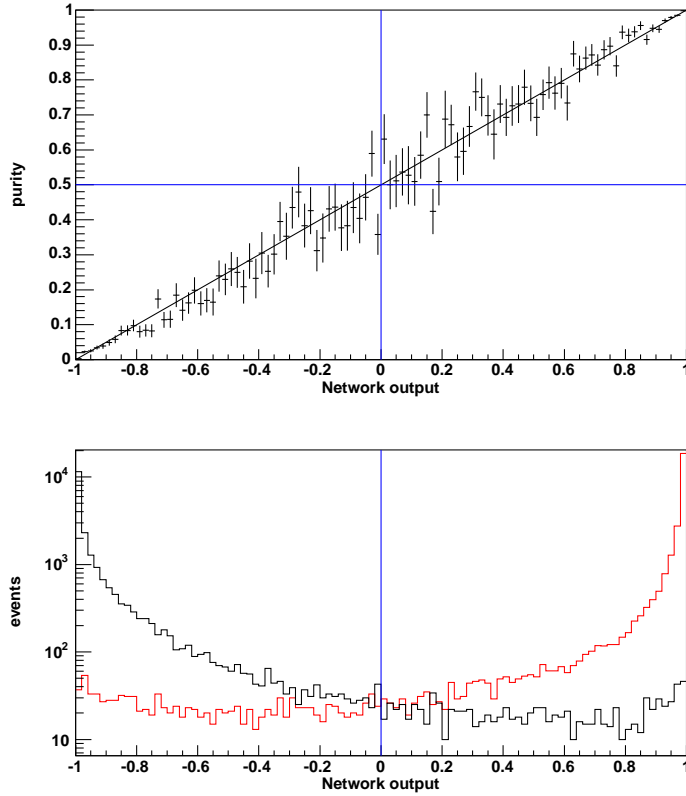


Figure B.2: Plots obtained after the NeuroBayes[®] training for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow \phi\pi, \phi \rightarrow KK$. The purity as a linear function of the network output (upper plot) and separation of signal and background (lower plot).

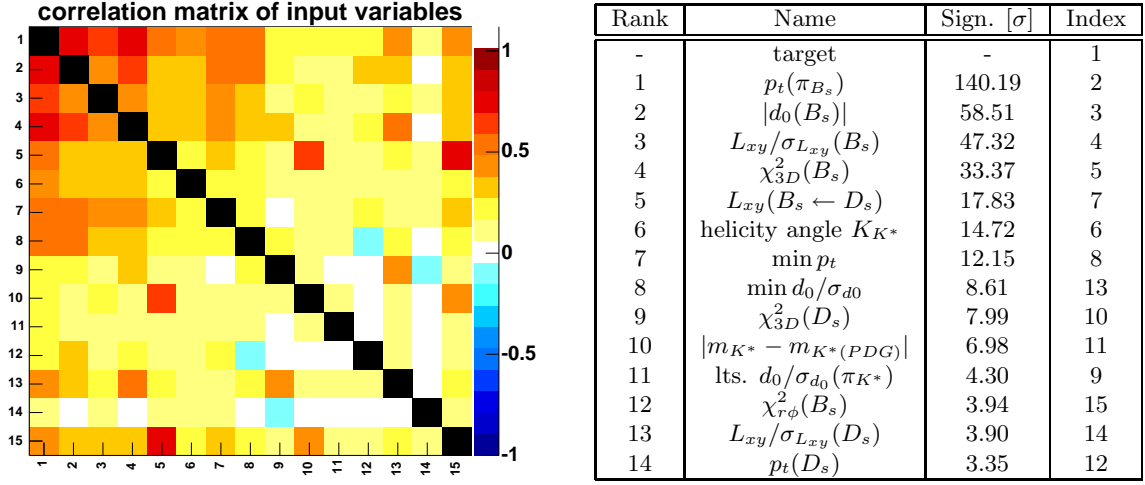


Figure B.3: Correlation matrix and significance table of the training variables for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K, K^* \rightarrow K\pi$.

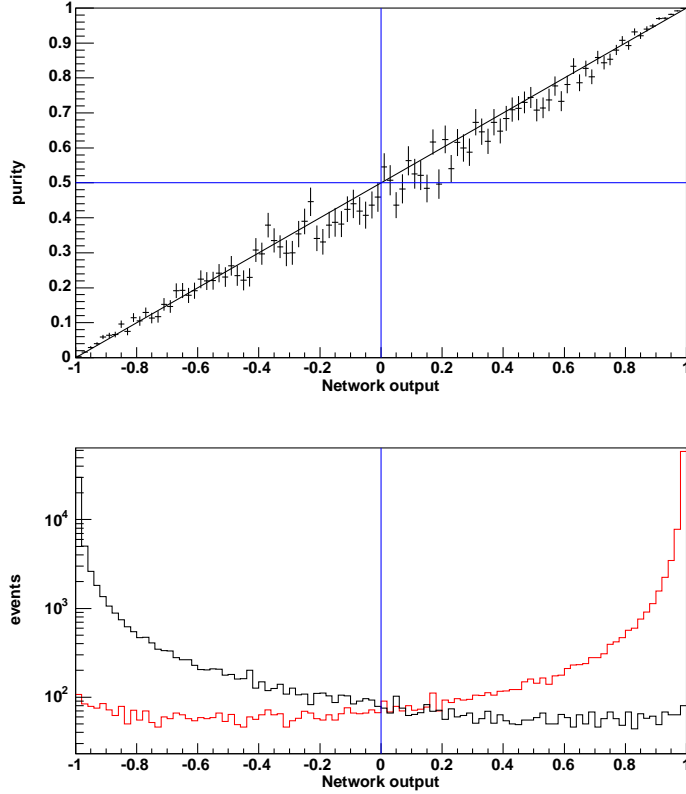


Figure B.4: Plots obtained after the NeuroBayes[®] training for the decay $B_s \rightarrow D_s\pi, D_s \rightarrow K^*K, K^* \rightarrow K\pi$. The purity as a linear function of the network output (upper plot) and separation of signal and background (lower plot).

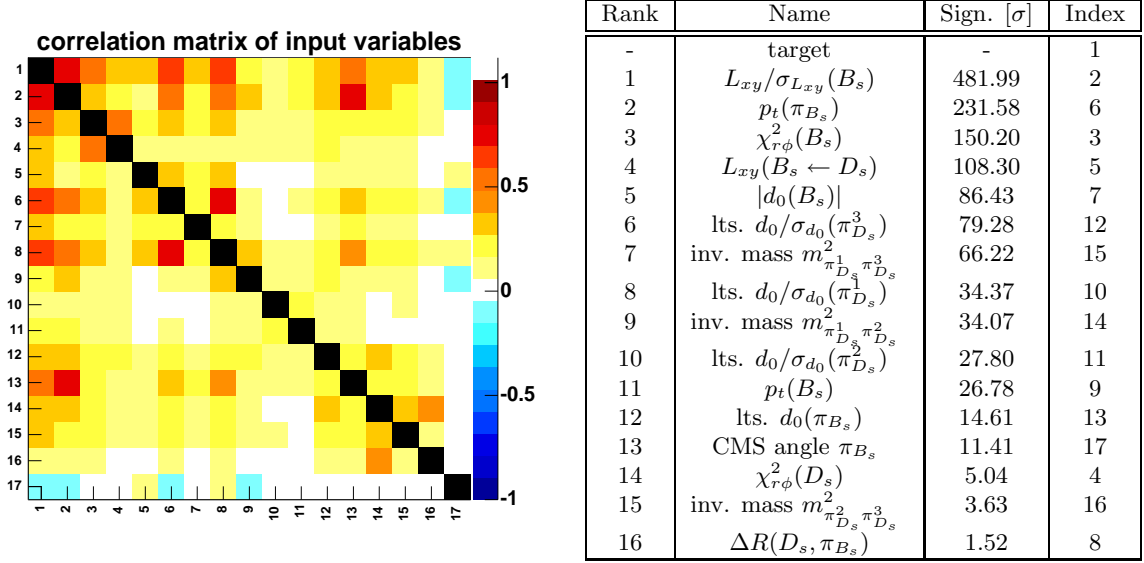


Figure B.5: Correlation matrix and significance table of the training variables for the decay $B_s \rightarrow D_s \pi, D_s \rightarrow \pi \pi \pi$.

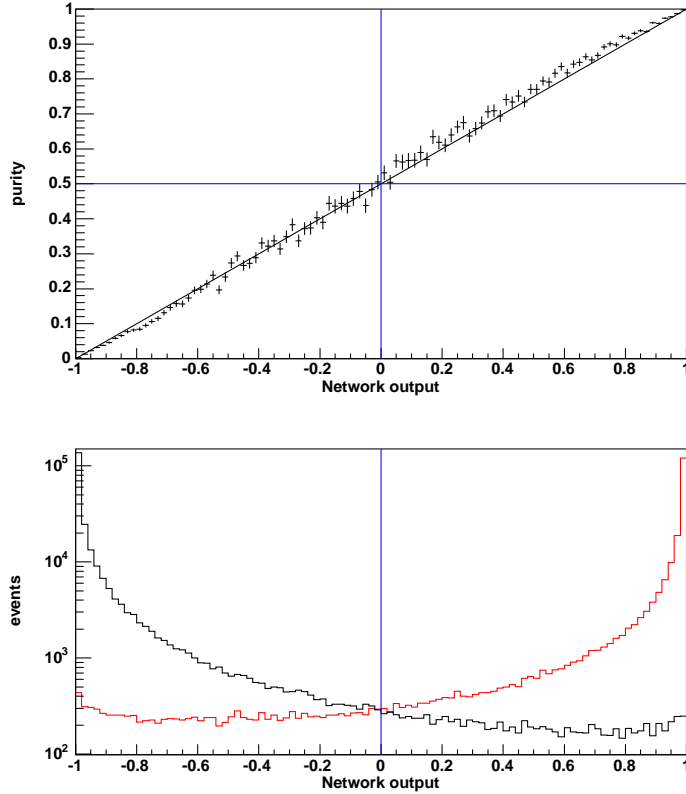


Figure B.6: Plots obtained after the NeuroBayes[®] training for the decay $B_s \rightarrow D_s \pi, D_s \rightarrow \pi \pi \pi$. The purity as a linear function of the network output (upper plot) and separation of signal and background (lower plot).

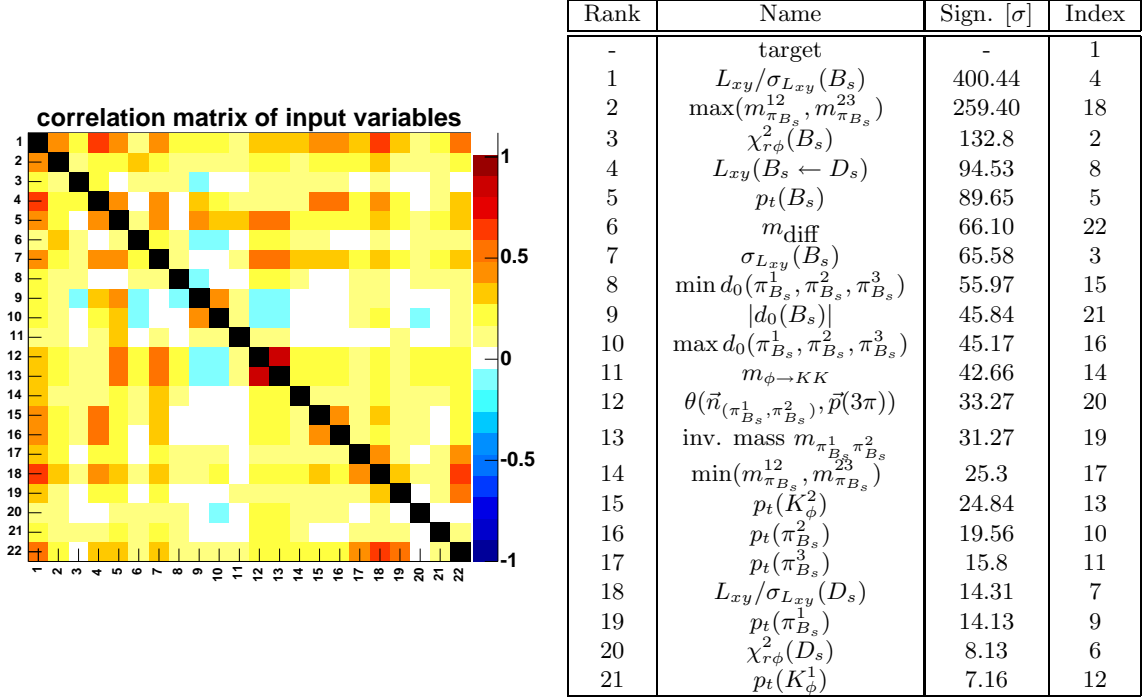


Figure B.7: Correlation matrix and significance table of the training variables for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow \phi\pi, \phi \rightarrow KK$.

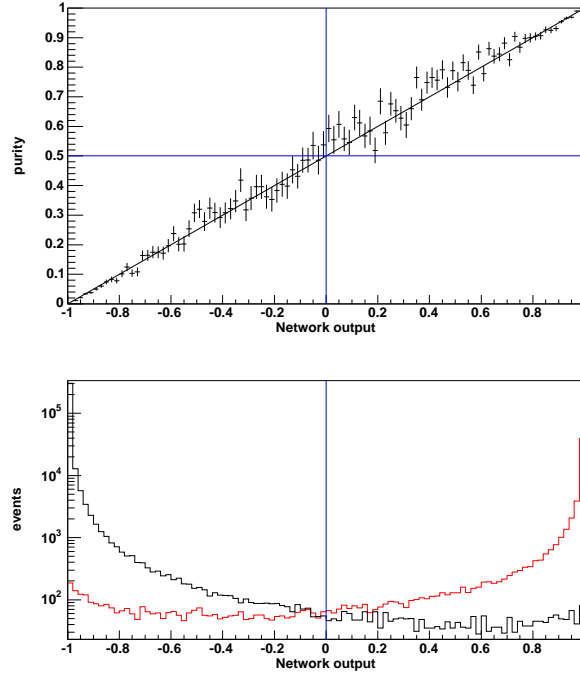


Figure B.8: Plots obtained after the NeuroBayes[®] training for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow \phi\pi, \phi \rightarrow KK$. The purity as a linear function of the network output (upper plot) and separation of signal and background (lower plot).

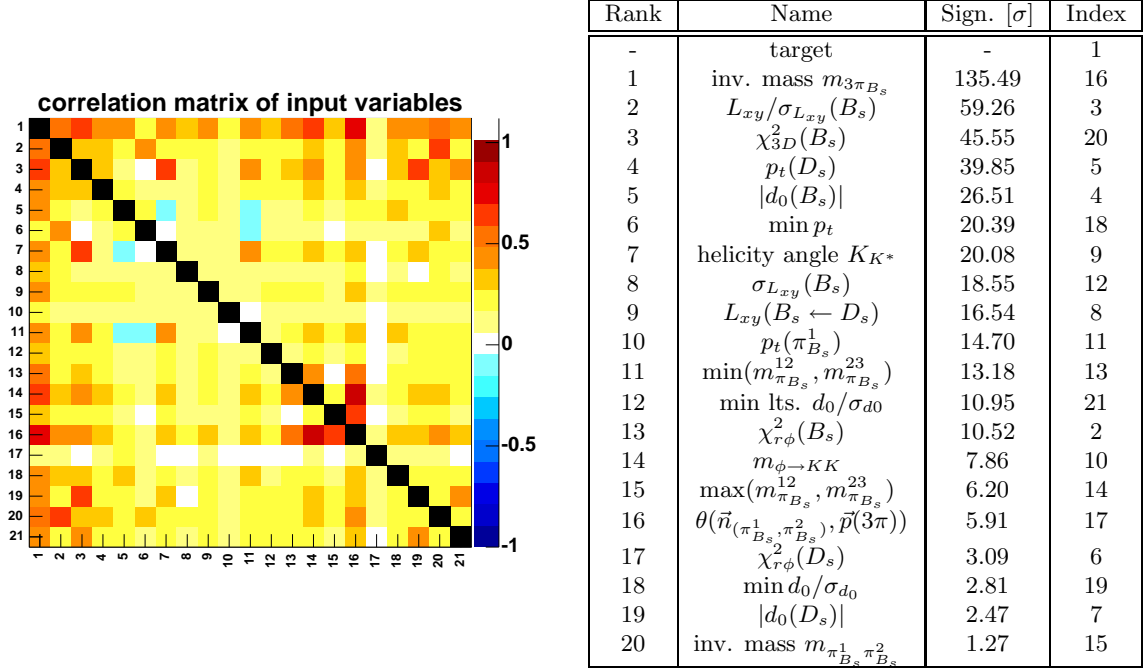


Figure B.9: Correlation matrix and significance table of the training variables for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow K^* K, K^* \rightarrow K\pi$.

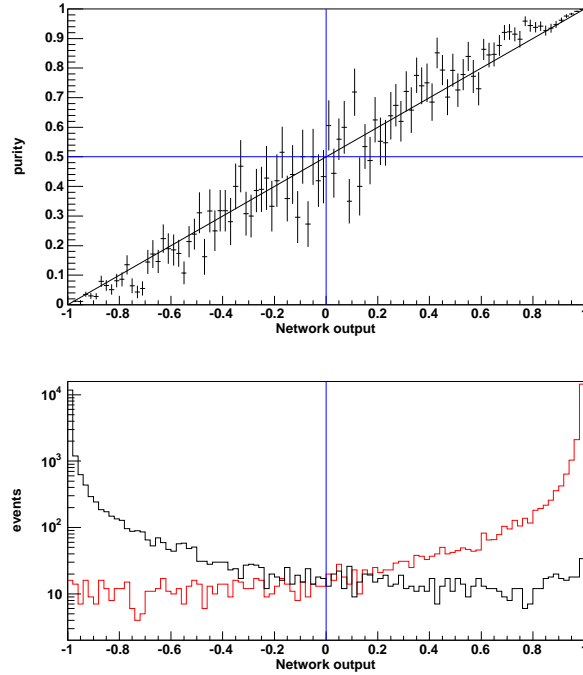


Figure B.10: Plots obtained after the NeuroBayes[®] training for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow K^* K, K^* \rightarrow K\pi$. The purity as a linear function of the network output (upper plot) and separation of signal and background (lower plot).

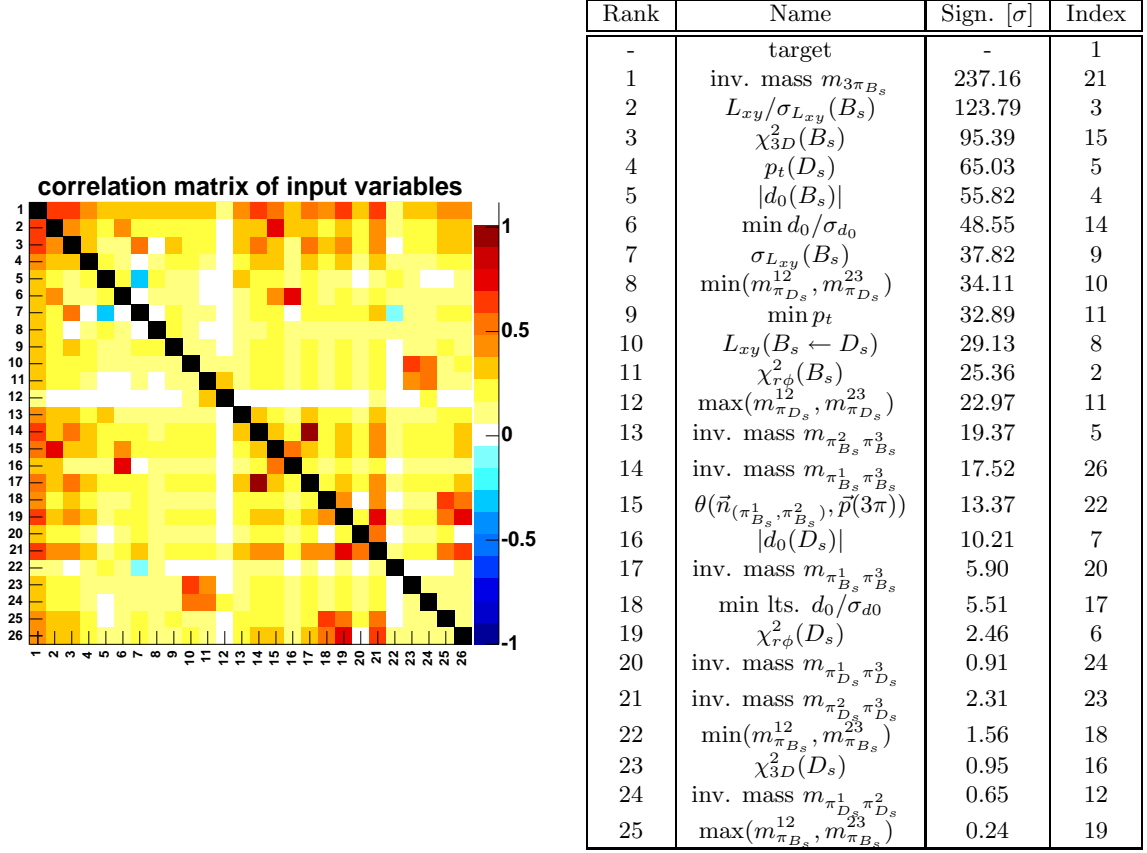


Figure B.11: Correlation matrix and significance table of the training variables for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow \pi\pi\pi$.

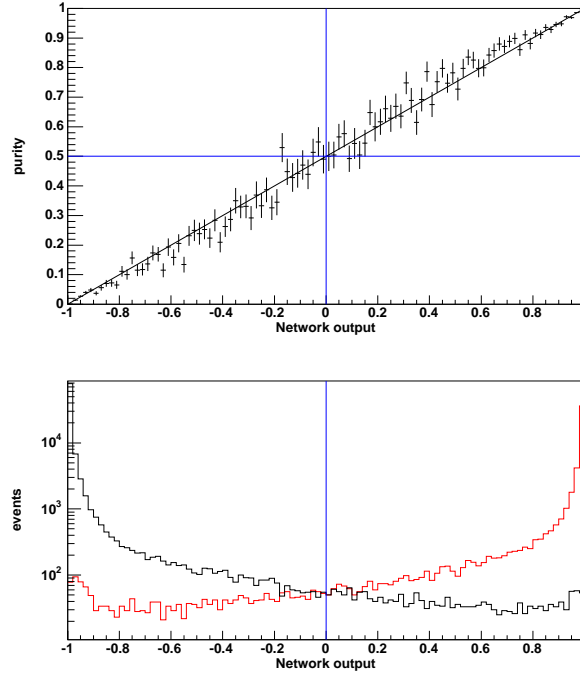


Figure B.12: Plots obtained after the NeuroBayes[®] training for the decay $B_s \rightarrow D_s 3\pi, D_s \rightarrow \pi\pi\pi$. The purity as a linear function of the network output (upper plot) and separation of signal and background (lower plot).

Appendix C

Compilation of Fit Templates

C.1 Fit Templates in Mass Space

C.1.1 Cabibbo suppressed signal

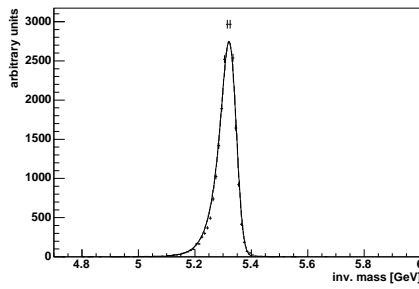


Figure C.1: MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$

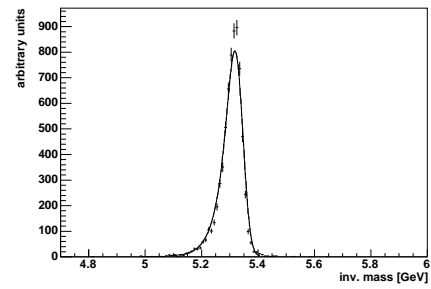


Figure C.2: MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

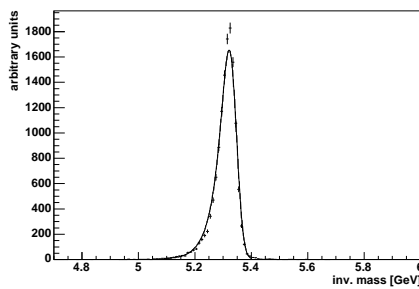


Figure C.3: MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$

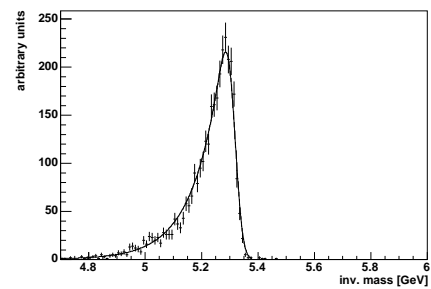


Figure C.4: MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi \pi$

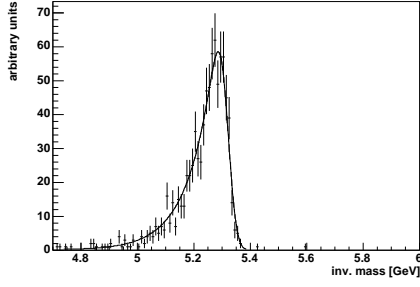


Figure C.5: MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$

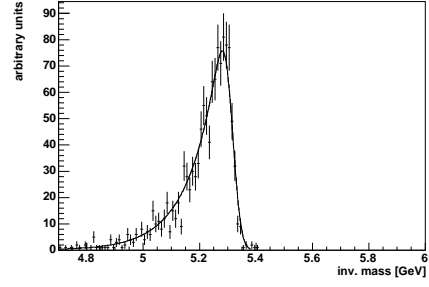


Figure C.6: MC fit template for the cabibbo suppressed decay of $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$

C.1.2 $D_s X$ background

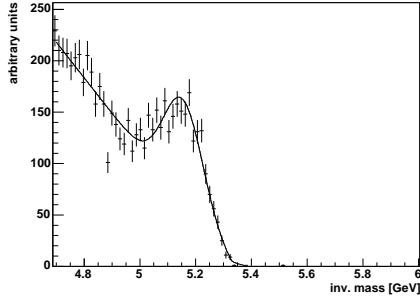


Figure C.7: MC fit template for the $D_s X$ background contribution of the $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$ decay mode

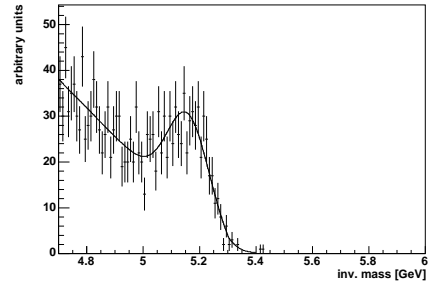


Figure C.8: MC fit template for the $D_s X$ background contribution of the $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ decay mode

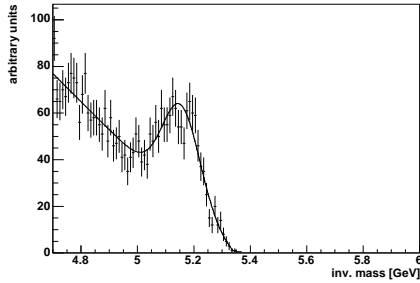


Figure C.9: MC fit template for the $D_s X$ background contribution of the $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$ decay mode

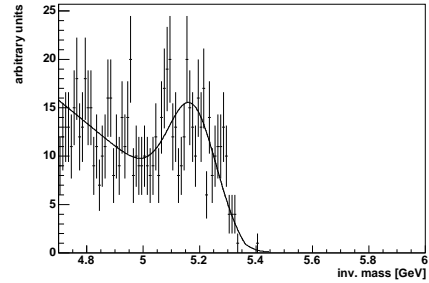


Figure C.10: MC fit template for the $D_s X$ background contribution of the $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi \pi$ decay mode

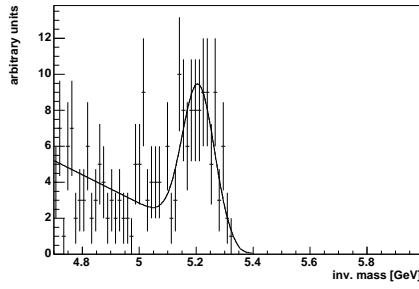


Figure C.11: MC fit template for the $D_s X$ background contribution of the $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$ decay mode

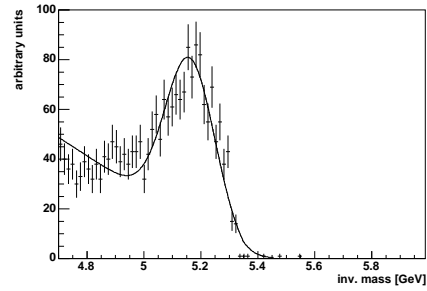


Figure C.12: MC fit template for the $D_s X$ background contribution of the $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$ decay mode

C.1.3 Λ_b background

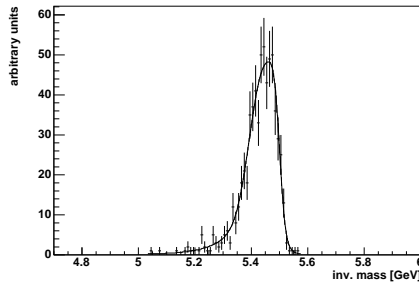


Figure C.13: MC fit template for the Λ_b contribution of the $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ decay mode

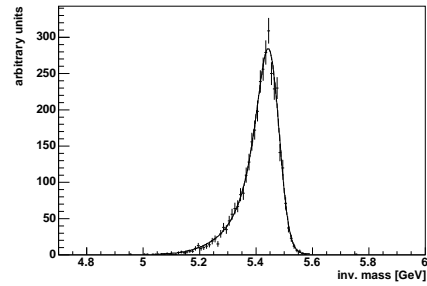


Figure C.14: MC fit template for the Λ_b contribution of the $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$ decay mode

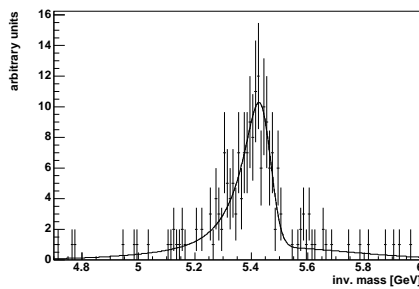


Figure C.15: MC fit template for the Λ_b contribution of the $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$ decay mode

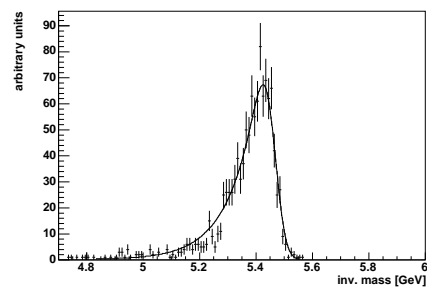


Figure C.16: MC fit template for the Λ_b contribution of the $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$ decay mode

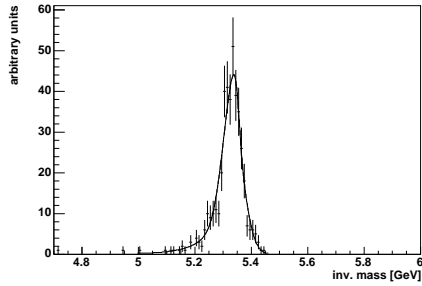
C.1.4 B^0 background

Figure C.17: MC fit template for the B^0 contribution of the $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ decay mode

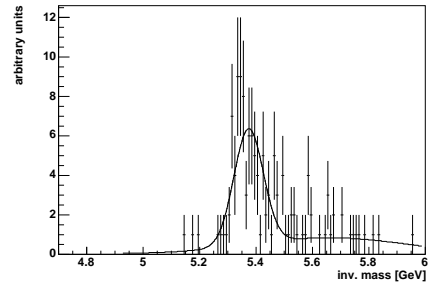


Figure C.18: MC fit template for the B^0 contribution of the $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$ decay mode

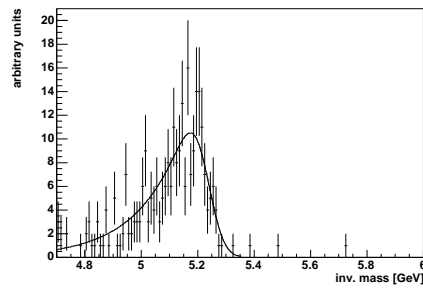


Figure C.19: MC fit template for the B^0 contribution of the $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$ decay mode

C.2 Fit Templates in Lifetime Space

C.2.1 Efficiency Function

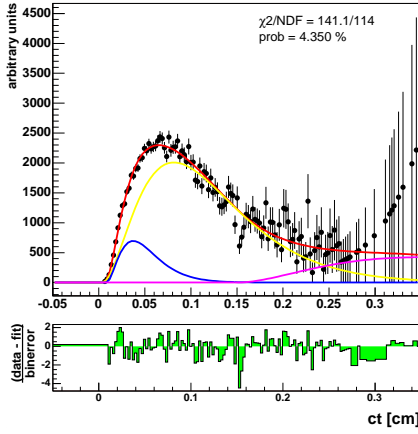


Figure C.20: Parameterized fit to the efficiency function for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$

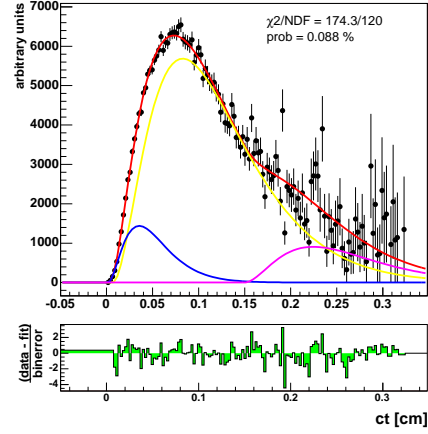


Figure C.21: Parameterized fit to the efficiency function for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

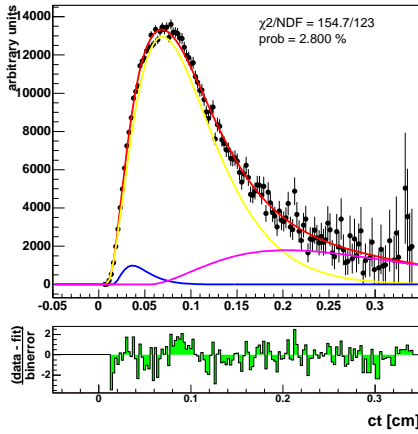


Figure C.22: Parameterized fit to the efficiency function for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$

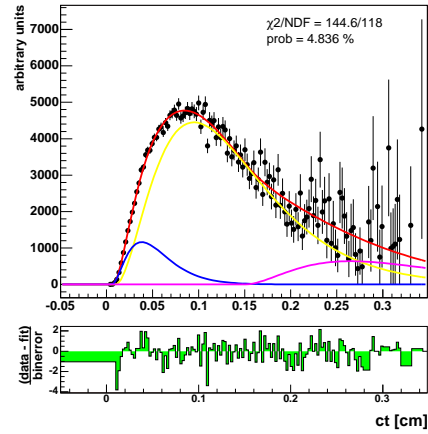


Figure C.23: Parameterized fit to the efficiency function for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi \pi$

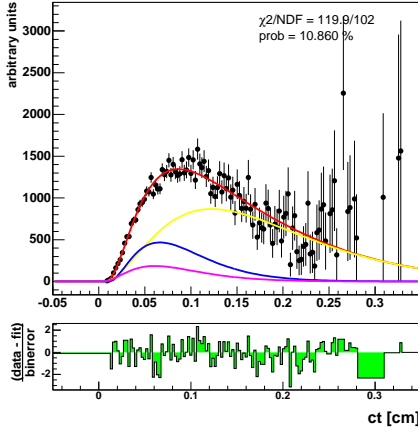


Figure C.24: Parameterized fit to the efficiency function for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$

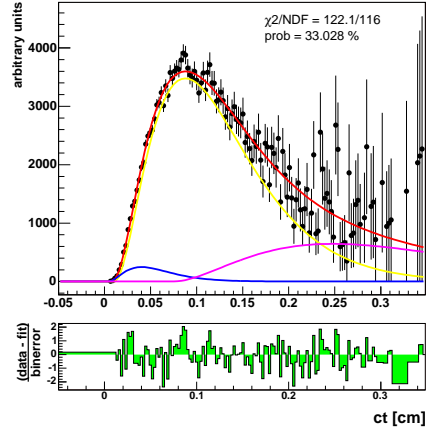


Figure C.25: Parameterized fit to the efficiency function for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$

C.2.2 Background Contributions to the Proper Decay Time

Combinatorial Background

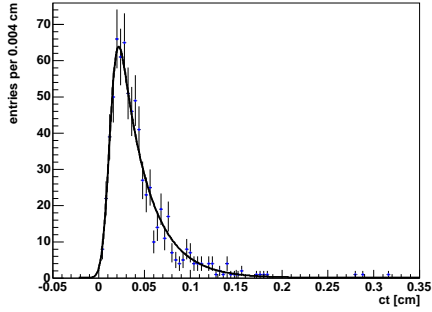


Figure C.26: Fit to the ct distribution of the upper sideband for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$

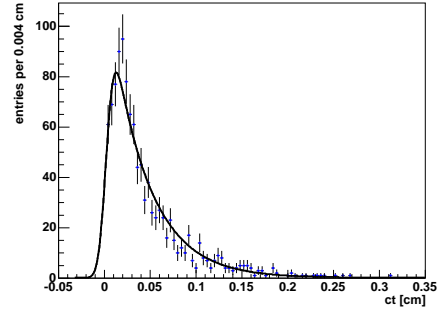


Figure C.27: Fit to the ct distribution of the upper sideband for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

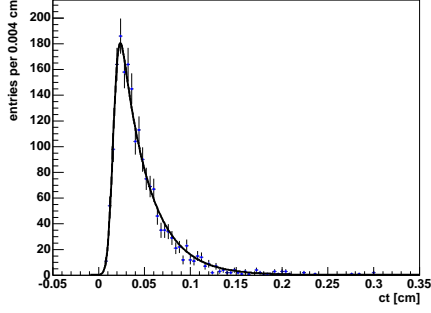


Figure C.28: Fit to the ct distribution of the upper sideband for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$

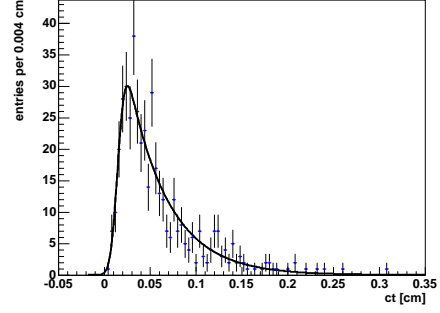


Figure C.29: Fit to the ct distribution of the upper sideband for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi\pi$

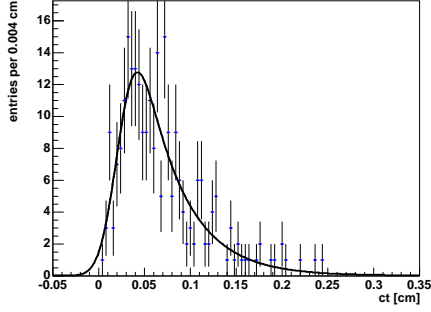


Figure C.30: Fit to the ct distribution of the upper sideband for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$

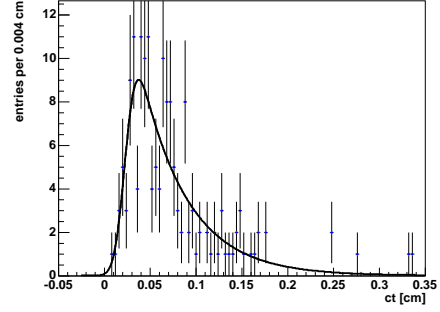


Figure C.31: Fit to the ct distribution of the upper sideband for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$

Λ_b Background

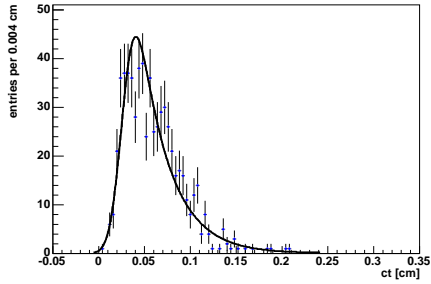


Figure C.32: Fit to the ct distribution of a Λ_b MC sample for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

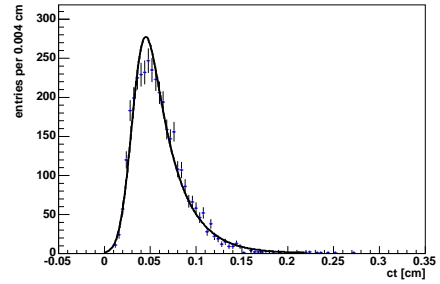


Figure C.33: Fit to the ct distribution of a Λ_b MC sample for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$

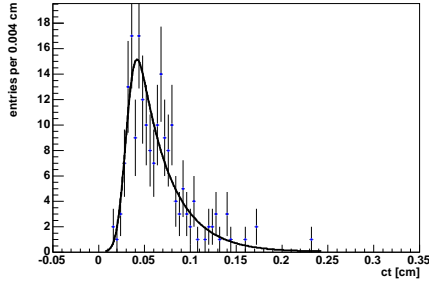


Figure C.34: Fit to the ct distribution of a Λ_b MC sample for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$

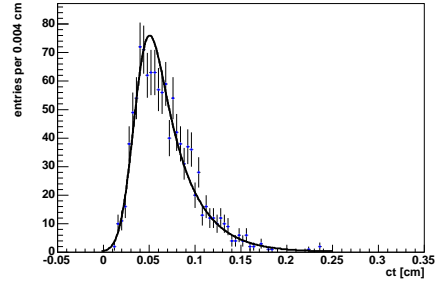


Figure C.35: Fit to the ct distribution of a Λ_b MC sample for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$

B^0 Background

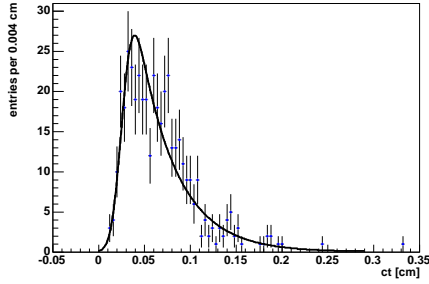


Figure C.36: Fit to the ct distribution of a B^0 MC sample for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

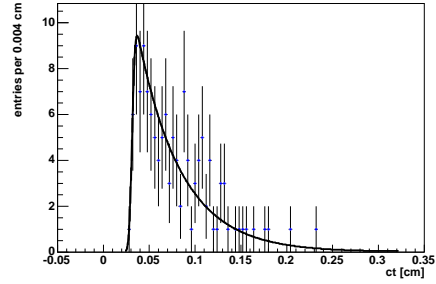


Figure C.37: Fit to the ct distribution of a B^0 MC sample for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^* K$

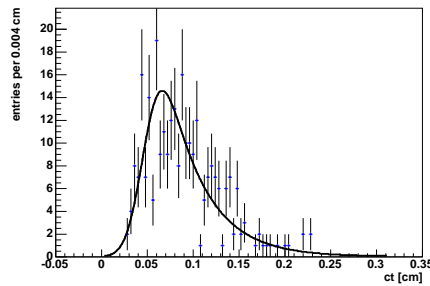


Figure C.38: Fit to the ct distribution of a B^0 MC sample for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$

Appendix D

Compilation of Fit Results

D.1 Binned Fits

D.1.1 Mass Fits

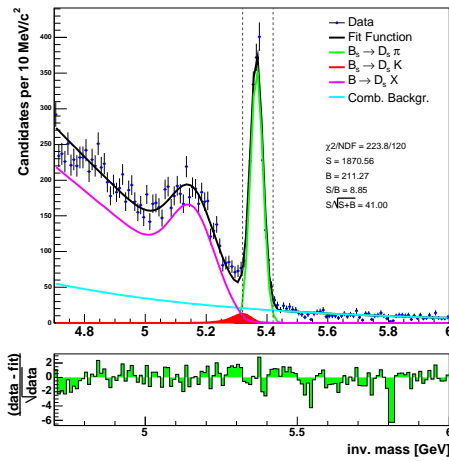


Figure D.1: Invariant mass spectrum for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$ and the binned fit result for a cut on the neural network output, $\text{nnout} > 0.68$.

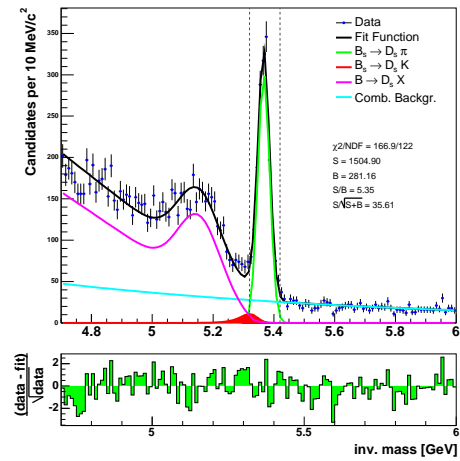


Figure D.2: Invariant mass spectrum for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$ and the binned fit result for a cut based analysis.

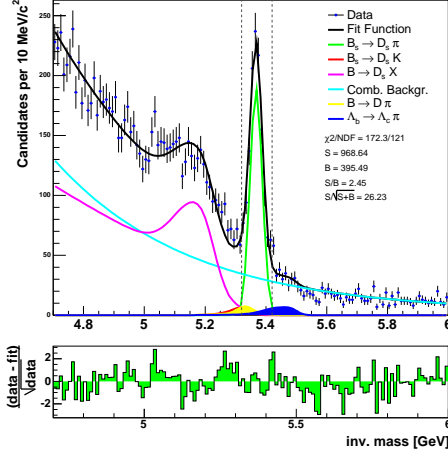


Figure D.3: Invariant mass spectrum for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ and the binned fit result for a cut on the neural network output, $\text{nnout} > 0.92$.

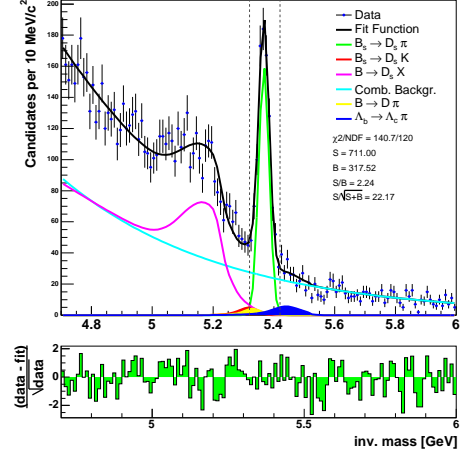


Figure D.4: Invariant mass spectrum for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$ and the binned fit result for a cut based analysis.

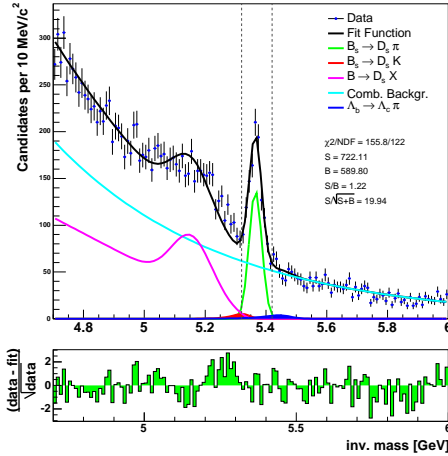


Figure D.5: Inv. mass spectrum for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$ and the binned fit result for a cut on the neural network output, $\text{nnout} > 0.48$.

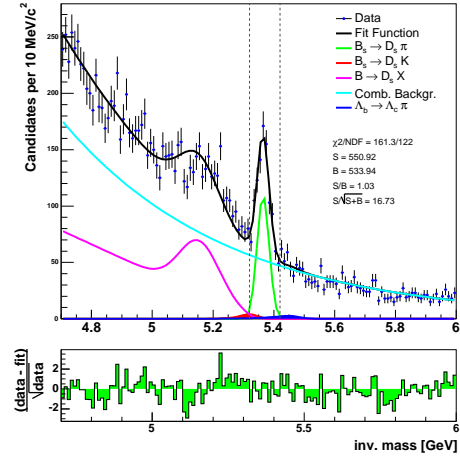


Figure D.6: Invariant mass spectrum for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$ and the binned fit result for a cut based analysis.

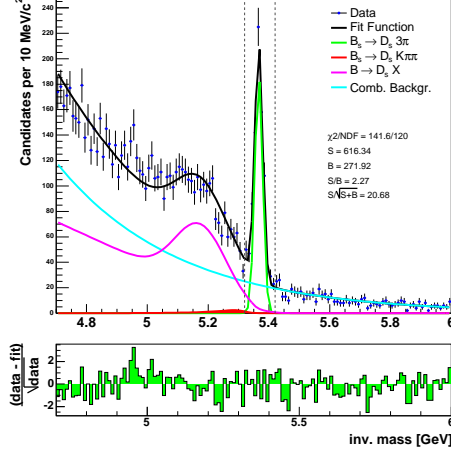


Figure D.7: Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi\pi$ and the binned fit result for a cut on the neural network output, $nnout > 0.90$.

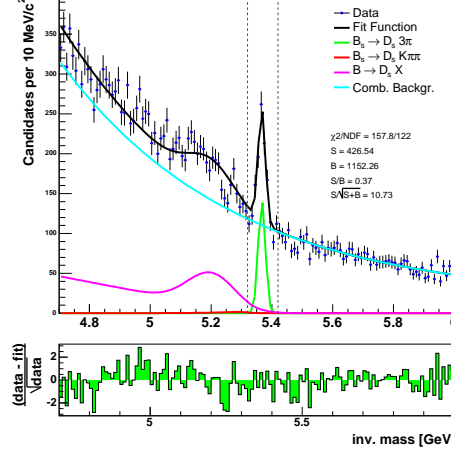


Figure D.8: Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi\pi$ and the binned fit result for a cut based analysis.

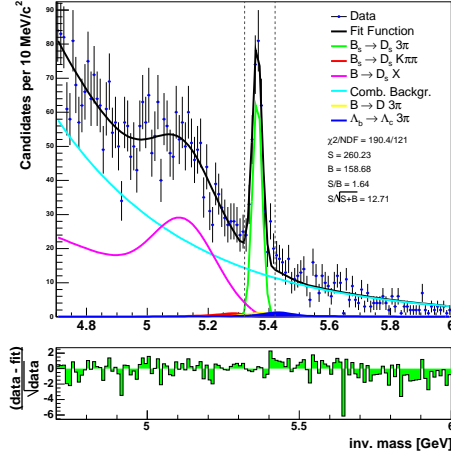


Figure D.9: Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^*K$ and the binned fit result for a cut on the neural network output, $nnout > 0.99$.

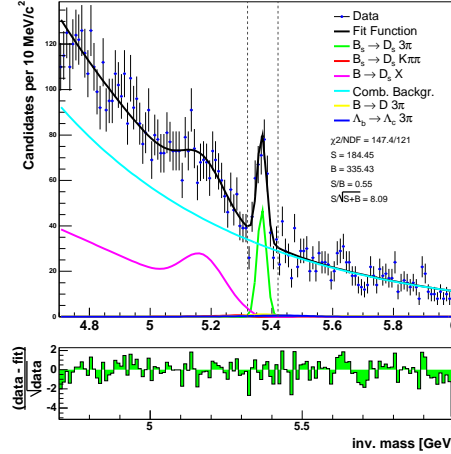


Figure D.10: Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^*K$ and the binned fit result for a cut based analysis.

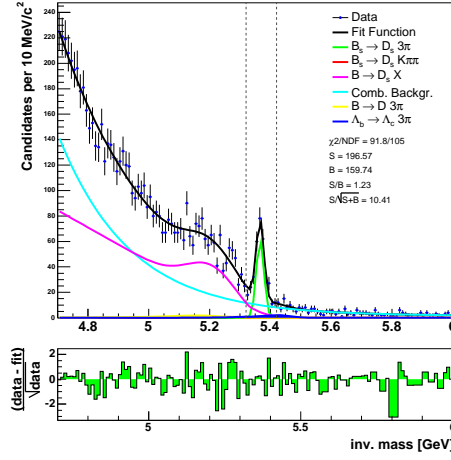


Figure D.11: Invariant mass spectrum for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$ and the binned fit result for a cut on the neural network output, $nnout > 0.99$.

D.2 Unbinned Fits

D.2.1 Projections of the Fit Results to the Mass and Lifetime Space

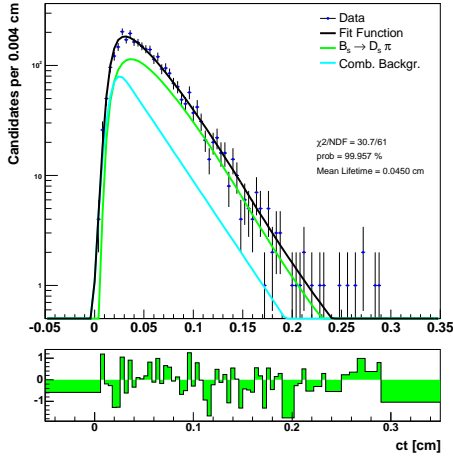


Figure D.12: Projection of the unbinned fit result to the lifetime space for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$

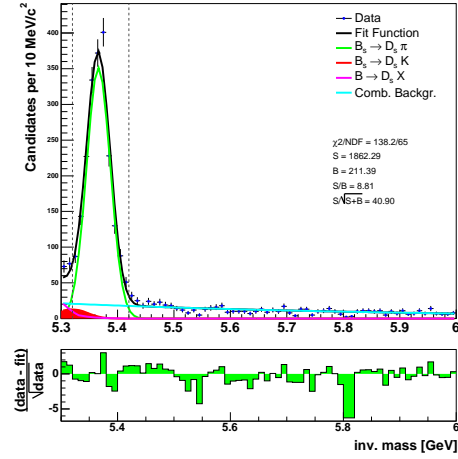


Figure D.13: Projection of the unbinned fit result to the mass space for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow \phi \pi$

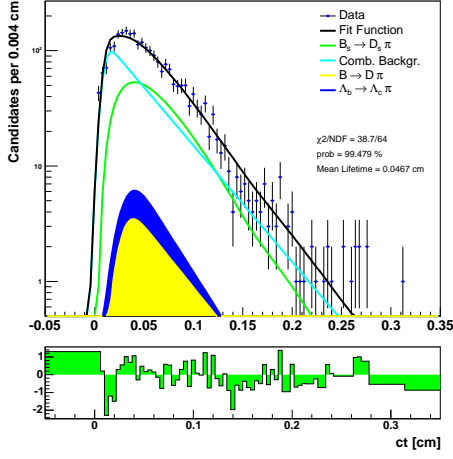


Figure D.14: Projection of the unbinned fit result to the lifetime space for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

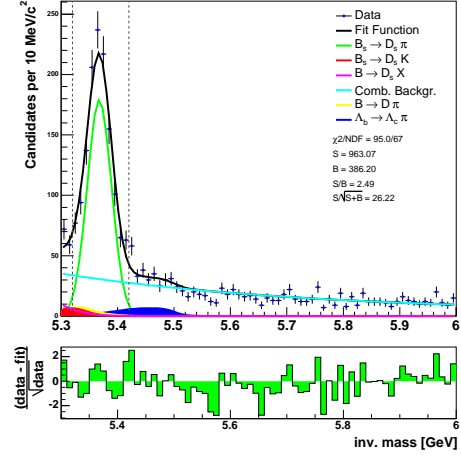


Figure D.15: Projection of the unbinned fit result to the mass space for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow K^* K$

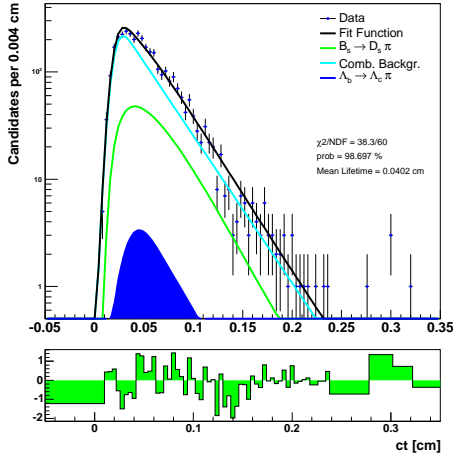


Figure D.16: Projection of the unbinned fit result to the lifetime space for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$

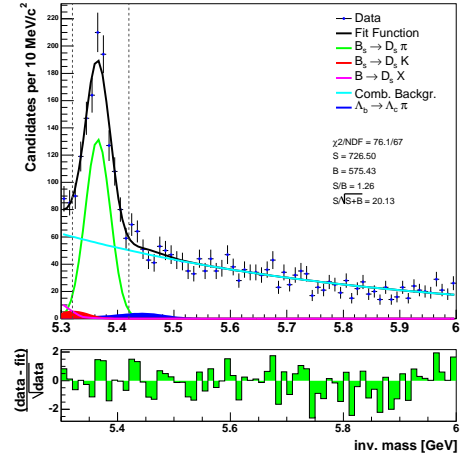


Figure D.17: Projection of the unbinned fit result to the mass space for the decay $B_s \rightarrow D_s \pi$, $D_s \rightarrow 3\pi$

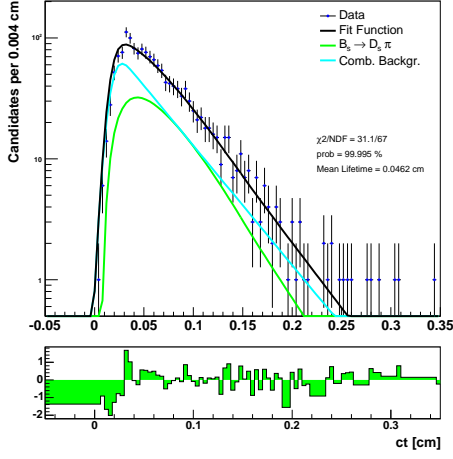


Figure D.18: Projection of the unbinned fit result to the lifetime space for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi\pi$

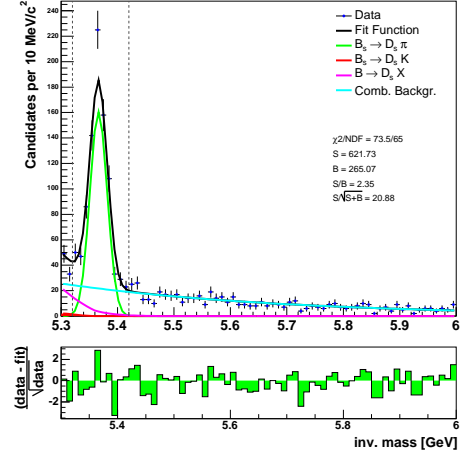


Figure D.19: Projection of the unbinned fit result to the mass space for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow \phi\pi$

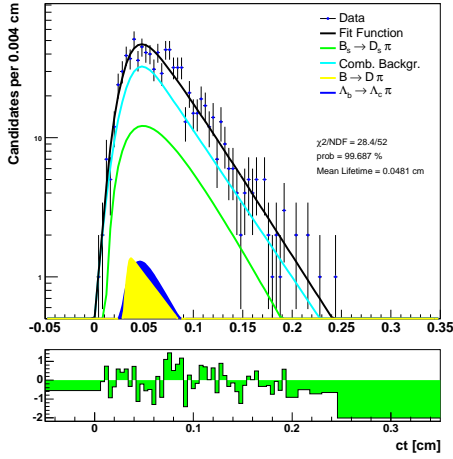


Figure D.20: Projection of the unbinned fit result to the lifetime space for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^*K$

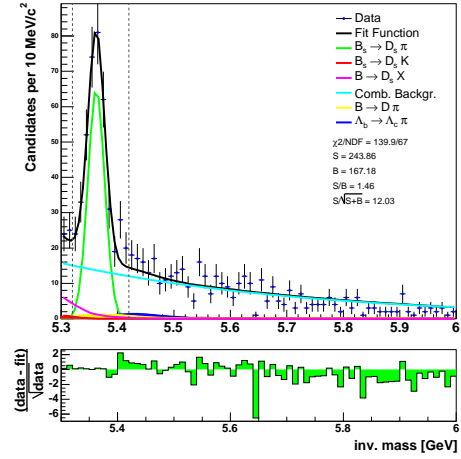


Figure D.21: Projection of the unbinned fit result to the mass space for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow K^*K$

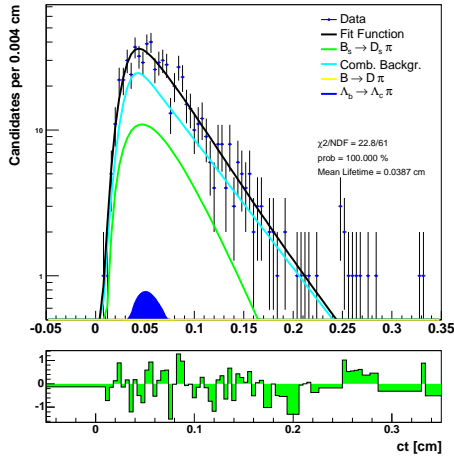


Figure D.22: Projection of the unbinned fit result to the lifetime space for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$

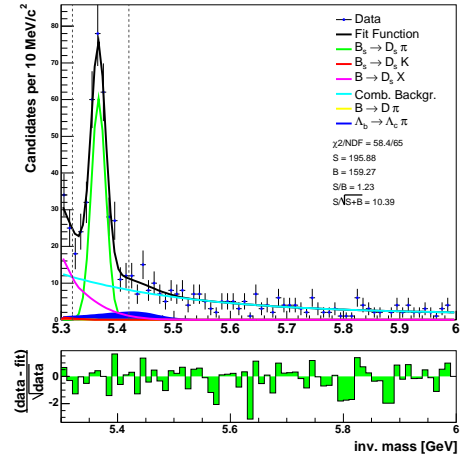


Figure D.23: Projection of the unbinned fit result to the mass space for the decay $B_s \rightarrow D_s 3\pi$, $D_s \rightarrow 3\pi$

Bibliography

- [1] The KTeV Collaboration. Measurements of direct CP violation, CPT symmetry, and other parameters in the neutral kaon system. *Phys. Rev.*, D67:012005, 2003.
- [2] The BABAR Collaboration. Measurement of the \bar{B}^0 lifetime and the $B^0\bar{B}^0$ oscillation frequency using partially reconstructed $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$ decays. *Phys. Rev.*, D73:012004, 2006.
- [3] The BELLE Collaboration. Improved measurement of CP-violation parameters $\sin 2\phi_1$ and $|\lambda|$, B meson lifetimes, and $B^0\text{-}\bar{B}^0$ mixing parameter Δm_d . *Phys. Rev.*, D71:072003, 2005.
- [4] The CDF Collaboration. Observation of $B_s^0\text{-}\bar{B}_s^0$ Oscillations. *Phys.Rev.Lett.*, 97:242003, 2006.
- [5] C. Dörr. Optimization of the signal selection of exclusively reconstructed B0 and Bs mesons at CDF-II. *PhD thesis*, IEKP-KA/2006-6, 2006.
- [6] A. Gessler. Rekonstruktion der Zerfälle $B_s \rightarrow D_s\pi$ und $B_s \rightarrow D_s\pi\pi\pi$ mit dem CDF-II-Detektor. *Diploma thesis*, IEKP-KA/2006-12, 2006.
- [7] Martin Heck, Michael Feindt, Michal Kreps (A. Abulencia et al., The CDF Collaboration). Observation of Orbitally Excited Bs Mesons. *CDF/PUB/BOTTOM/CDFR/8712*, 2007.
- [8] The BABAR Collaboration. Evidence for $D^0\text{-}\bar{D}^0$ Mixing at Babar. *arXiv:0705.2785 [hep-ex]*, 2007.
- [9] The BELLE Collaboration. Evidence for $D^0\text{-}\bar{D}^0$ Mixing. *Phys. Rev. Lett.*, 98:211803, 2007.
- [10] Bs Mixing Group (Contact: A. Belloni, I. Furic, G. Gomez-Ceballos, I. Kravchenko, N. Leonardo, S. Menzemer, J. Miles, Ch. Paus, J. Piedra). Same Side Kaon Tagging in Fully Reconstructed Decays. *CDF/PUB/BOTTOM/PUBLIC/8206*, 2006.
- [11] A. Schimdt. Combined B Flavour Tagging with Artificial Neural Networks. *Diploma thesis*, IEKP-KA/2006-11, 2006.
- [12] S. L. Glashow. *Nucl. Phys.*, 22:579, 1961.
- [13] S. Weinberg. *Phys. Rev. Lett.*, 19:1264, 1967.
- [14] A. Salam. *in Elementary Particle Theory*, 1968.

- [15] D. J. Gross and F. Wilczek. *Phys. Rev.*, D8:3633–3652, 1973.
- [16] W.-M. Yao et al. Review of Particle Physics. *Journal of Physics G33,1*, 2006.
- [17] J. Hosaka et al. Solar neutrino measurements in Super-Kamiokande-I. *Phys. Rev.*, D73:112001, 2006.
- [18] S. Yasumi et al. The mass of the electron-neutrino from electron capture in Ho-163. *Phys. Lett.*, B334:229–233, 1994.
- [19] K. Assamagan et al. Upper limit of the muon-neutrino mass and charged pion mass from momentum analysis of a surface muon beam. *Phys. Rev.*, D53:6065–6077, 1996.
- [20] R. Barate et al. An upper limit on the tau neutrino mass from three- and five-prong tau decays. *Eur. Phys. J.*, C2:395–406, 1998.
- [21] Ch. Kraus et al. Final results from phase II of the mainz neutrino mass search in tritium beta decay. *Eur. Phys. J.*, C40:447–468, 2005.
- [22] S. Khalil and E. Torrente-Lujan. Neutrino mass and oscillation as probes of physics beyond the standard model. *J. Egyptian Math. Soc.*, 9:91–141, 2001.
- [23] P. Schmüser. *Feynman-Graphen und Eichtheorien für Experimentalphysiker*. Springer-Verlag, 1988.
- [24] A. Pich. Weak decays, quark mixing and CP violation: Theory overview. *Nucl. Phys. Proc. Suppl.*, 66:456–465, 1998.
- [25] N. Cabibbo. Unitary Symmetry And Leptonic Decays. *Phys. Rev. Lett.*, 10:531, 1963.
- [26] T. Maskawa M. Kobayashi. CP-violation in the renormalizable theory of weak interaction. *Progr. Theor. Phys.*, 49:652–657, 1973.
- [27] F. J. Gilman et al. *Phys. Lett.*, B592:793, 2004.
- [28] Wolfenstein, Lincoln . Parametrization of the Kobayashi-Maskawa Matrix. *Phys. Rev. Lett.*, 51(21):1945–1947, 1983.
- [29] J. Charles et al. [CKM Fitter Group]. CP Violation and the CKM Matrix: Assessing the Impact of the Asymmetric B Factories. *Eur. Phys. J.*, C41:1–131, 2005.
- [30] A. Höcker et. al. A New Approach to a Global Fit of the CKM Matrix. *Eur. Phys. J.*, C21:225–259, 2001.
- [31] C. N. Yang T.T. Wu. *Phys. Rev. Lett.* 13,380, 1964.
- [32] Michel Beyer. *CP Violation in Particle, Nuclear and Astrophysics*. Springer-Verlag, 2002.
- [33] E. P. Wigner V. F. Weisskopf. Berechnung der natürlichen Linienbreite auf Grund der Diracschen Lichttheorie. *Zeitschrift für Physik*, 63:54–73, 1930.
- [34] P. H. Weisz A. J. Buras, M. Jamin. *Nucl. Phys.*, B245, 1984.

- [35] C. S. Lim T. Inami. *Progr. Theor. Phys.*, 65:297, 1981.
- [36] P. H. Weisz A. J. Buras, M. Jamin. *Nucl. Phys.*, B347:491, 1990.
- [37] M. Okamoto. Full determination of the CKM matrix using recent results from lattice QCD. *hep-lat/0510113*, 2005.
- [38] W.-M. Yao et al. *J. Phys. G33,1*, 2006.
- [39] The CDF Collaboration. Measurement of the Bottom-Quark Hadron Masses in Exclusive J/ψ Decays with the CDF Detector. *Phys. Rev. Lett.*, 96:202001, 2006.
- [40] Y. Nir. *Phys. Lett.*, B327:85, 1994.
- [41] M. Benecke et al. Next-to-leading order QCD corrections to the lifetime difference of B_s mesons. *Phys. Lett.*, B459:631–640, 1999.
- [42] FNAL Accelerator Division. Fermilab’s Chain of Accelerators. <http://www-bd.fnal.gov/public/chain.html>.
- [43] FNAL Accelerator Division. Run II Parameter List. <http://www-ad.fnal.gov/runII/index.html>.
- [44] CDF ACE Help. Tevatron Rookie Book. http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html.
- [45] The CDF II Collaboration. The CDF II Detector Technical Design Report. *FERMILAB-Pub-96/390-E*, 1996.
- [46] CDF ACE Help. An Ace’s Guide to the CDF Run II Detector. http://www-cdfonline.fnal.gov/ops/ace2help/detector_guide/detector_guide.html.
- [47] Kevin Burkett. Design and Construction of the CDF Central Outer Tracker. *CDF/PUB/TRACKING/PUBLIC/5321*, 2000.
- [48] A. Mukherjee and R. Wagner. COT Central Outer Tracker. *CDF/PUB/TRACKING/PUBLIC/6267*, 2003.
- [49] Philipp Mack. Kaonidentifikation im Silizium-Vertex-Detektor bei CDF Run II. *Diploma thesis*, IEKP-KA/2003-24, 2003.
- [50] H. Frisch et al. Conceptual Design of a Deadtimeless Trigger for the CDF Trigger upgrade. *CDF/DOC/TRIGGER/CDFR/2038*, 1994.
- [51] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society, Series A*, 222:309–368, 1922.
- [52] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Teubner Studienbücher Physik, 1998.
- [53] W. T. Eadie et al. *Statistical Methods in Experimental Physics*. North Holland Publishing Company, 1971.

- [54] C. G. Broydon. The convergence of a class of double rank minimization. *Inst. Maths. Applics.*, 6:222–231, 1970.
- [55] R. Fletcher. A new approach to variable metric algorithms. *Comp. J.*, 13(3):317–322, 1970.
- [56] D. Goldfarb. A family of variable metric methods derived by variational means. *Mathematics of Computation*, 24:23–26, 1970.
- [57] D. F. Shanno. Conditions of quasi-newton methods for function minimization. *Mathematics of Computation*, 24:647–656, 1970.
- [58] *ROOT User's Guide v5.14*, 2006.
- [59] F. James and M. Roos. Minuit - a system for function minimization and analysis of parameter errors and correlations. *Comp. Phys. Comm.*, 10:343–367, 1972.
- [60] M. Feindt. A Neural Bayesian Estimator for Conditional Probability Densities. *arXiv:physics/0402093v1*, 2004.
- [61] The late Rev. Mr. T. Bayes. An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.
- [62] The Data Quality Monitoring Group. GOOD RUN LIST Page. <http://www-cdf.fnal.gov/internal/dqm/goodrun/good.html>.
- [63] T. Scheidle. Forward Tracking with the silicon vertex detector at the CDF experiment in RUN II. *PhD thesis*, IEKP-KA/2007-5, 2007.
- [64] J. Boudreau, C. Green, P. Murat, M. Paterno, L. Sexton-Kennedy. The CDF Geometry System for Run II Offline Software. A Guide for Subdetector Package Writers. *CDF/DOC/COMP_UPG/PUBLIC/5368*, 2000.
- [65] K. Anikeev, G. Bauer, I. Furic, S. Gromoll, A. Korn, I. Kravchenko, Ch. Paus, A. Rakitin, J. Tseng. Calibration of Energy Loss and Magnetic Field using J/ψ Events in Run II. *CDF/DOC/BOTTOM/CDFR/5958*, 2002.
- [66] M. Feindt, S. Menzemer and K. Rinnert. TrackingKal - A Tracking and Alignment Software Package for the CDFII Silicon Detector. *CDF/THESIS/TRACKING/PUBLIC/5968*, 2003.
- [67] Ch. Paus et. al. <http://cdfkits.fnal.gov/CdfCode/source/BottomMods/>.
- [68] C. Blocker, J. Boudreau, Ch. Paus. Common Tools for B Physics Analyses in Run II at CDF. *CDF/DOC/BOTTOM/CDFR/5735*, 2002.
- [69] J. MARRINER. SECONDARY VERTEX FIT WITH MASS AND POINTING CONSTRAINTS (CTVMFT). *CDF/DOC/SEC_VTX/PUBLIC/1996*, 1993.
- [70] K. Anikeev, A. Belloni, G. Gomez-Ceballos, S. Menzemer, Ch. Paus, J. Piedra. Monte Carlo Samples for Δm_s Measurement. *CDF/DOC/BOTTOM/GROUP/8303*, 2006.

- [71] Torbjorn Sjostrand, Leif Lonnblad and Stephen Mrenna. PYTHIA 6.2: Physics and manual. *LU TP 01-21, hep-ph/0108264*, 2001.
- [72] E. Norrbin and T. Sjostrand. Production and hadronization of heavy quarks. *Eur. Phys. J.*, C17:137–161, 2000.
- [73] D. J. Lange. The evtgen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.
- [74] The CDF Detector Simulation. http://www-cdf.fnal.gov/cdfsims/cdfsims_main.html.
- [75] The trigger simulation project. <http://ncdf70.fnal.gov:8001/trigsim/trgsim.html>.
- [76] K. Anikeev, P. Murat, Ch. Paus. Description of Bgenerator II. *CDF/DOC/BOTTOM/CDFR/5092*, 1999.
- [77] Bs Mixing Group (Contact: I.Furic, G.Di Giovanni, G.Gomez-Ceballos, T.Kuhr, V. Tiwari). Combined opposite flavor tagger. *CDF/DOC/BOTTOM/PUBLIC/8460*, 2006.
- [78] T. Allmendinger, G. Barker, M. Feindt, C. Haag, M. Moch. BSAURUS - A Package For Inclusive B-Reconstruction in DELPHI. *DELPHI 2000-069 PHYS 868*, 2000.
- [79] The CDF Collaboration. Measurement of the Ratios of Branching Fractions $\mathcal{B}(B_s^0 \rightarrow D_s^- \pi^+ \pi^+ \pi^-) / \mathcal{B}(B^0 \rightarrow D^- \pi^+ \pi^+ \pi^-)$ and $\mathcal{B}(B_s^0 \rightarrow D_s^- \pi^+) / \mathcal{B}(B^0 \rightarrow D^- \pi^+)$. *CDF/PUB/BOTTOM/PUBLIC/8437*, 2006.
- [80] Bs Mixing group (Contact: A. Belloni, I.Furic, G. Gomez-Ceballos, I. Kravchenko, N. Leonardo, S. Menzemer, J. Miles, Ch. Paus, J.Piedra). Updated Study of Bs Oscillations in $B_s \rightarrow D_s(3)\pi$. *CDF/PHYS/BOTTOM/CDFR/7861*, 2005.
- [81] Bs Mixing Group (Contact: A. Belloni, I. Furic, G. Gomez-Ceballos, I. Kravchenko, N. Leonardo, S. Menzemer, J. Miles, Ch. Paus, J. Piedra). Scale Factors for Proper Time Uncertainties at CDF. *CDF/PUB/BOTTOM/PUBLIC/7944*, 2005.
- [82] Bs Mixing Group (Contact: A.Belloni, I.Furic, G. Di Giovanni, G.Gomez-Ceballos, A.Heijboer, I.Kravchenko, N.Leonardo, S.Menzemer, J.Miles, Ch.Paus, J.Piedra). Validation of New Datasets for the Bs Oscillation Study in $B_s \rightarrow D_s(3)\pi$ decays using OST and SSKT. *CDF/DOC/BOTTOM/CDFR/8172*, 2006.
- [83] Heavy Flavor Averaging Group. Averages of b-hadron properties at the end of 2006. *arXiv:0704.3575v1 [hep-ex]*, 2006.
- [84] J. Fränkle. Fitting Procedure for the Determination of the Bs Mixing Frequency. *Diploma thesis*, IEKP-KA/2007-3, 2007.

Acknowledgements

First of all I would like to thank Prof. Dr. Michael Feindt for supporting me throughout the thesis and helping whenever problems of physical or statistical nature occurred. I would also like to thank Prof. Dr. Günter Quast for the co-supervision of this thesis. I would like to thank Prof. Dr. Thomas Müller for the possibility of visiting the Fermilab site. The three month I spent there working directly at the experiment have been a exciting experience.

Furthermore I want to express my gratitude to Dr. Michal Kreps for helping to overcome the technical and physical difficulties of this analysis and proofreading this manuscript. I would like to thank Dr. Thomas Kuhr for always providing helpful information, especially about problems concerning the B flavor tagger, and for reading this thesis carefully.

A special thanks goes to the following people for their contribution to the final result of this thesis: Dr. Christian Dörr, Jan Fränkle, Andreas Gessler, Jan Morlock and Andreas Schmidt.

The financial support of this thesis was provided by Gradiuiertenkolleg Hochenergie and Teilchenastrophysik and the Landesgraduiertenkolleg Baden-Württemberg and I want to thank them for their scholarships.

Another thanks goes to the EKP Admin team and their continuous efforts of keeping all computational resources available. I want to thank all current and former roommates of our office 9/2 for the great ambiance making the daily work more enjoyable.

Additionally I would like to mention the people of the EKP who enriched my life beyond particle physics: Volker Büge, Joachim Heuser, Michael Milnik, Andreas Oehler, Svenja Richter, Thorsten Scheidle, Armin Scheurer and Alexander Schmidt.

Furthermore I want thank all other current and former members of the EKP for their support and the pleasant working environment.

Finally I would like to thank my parents for their continuous support, their patience and their confidence in me.