Ein neuronaler Jet-Ladungs-Algorithmus für die Unterscheidung von b- und $\bar{b}$-Quark-Jets in B_s^0-$\bar{B}_s^0$-Oszillationsfrequenz-Messungen

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

genehmigte

DISSERTATION

von

Dott.ssa Claudia Lecci
aus Poggiardo (Italien)

Tag der mündlichen Prüfung: 01.07.2005
Referent: Prof. Dr. M. Feindt, Institut für Experimentelle Kernphysik
Korreferent: Prof. Dr. G. Quast, Institut für Experimentelle Kernphysik

Deutsche Zusammenfassung

Einleitung

Der Übergang neutraler Mesonen zwischen Teilchen- und Anti-Teilchen-Zustand wird *Oszillation* (oder *Mischung*) genannt. Dieses Phänomen ist im Rahmen des Standardmodells dadurch erklärt, dass die Eigenzustände der schwachen Wechselwirkung nicht die Masseneigenzustände sind.
Die Mischung wurde 1955 für das K^0-$\bar{K}^0$-System vorausgesagt und 1956 zum ersten Mal beobachtet. Die Mischung von neutralen B-Mesonen wurde das erste Mal 1987 im B_d^0-System nachgewiesen. Die Oszillationsfrequenz des B_d^0-Mesons, Δm_d, wurde von verschiedenen Experimenten genau gemessen und der derzeitige Weltmittelwert ist $\Delta m_d = 0.502 \pm 0.007$ ps^{-1}.
Die Oszillation des B_s^0-Systems wurde beobachtet, die Frequenz Δm_s aber noch nicht bestimmt. Das Standardmodell sagt voraus, dass B_s^0-Oszillationen sehr schnell sind, und dass der Wert von Δm_s in der Größenordnung von 10 ps^{-1} liegt. Ein Wert größer als 25 ps^{-1} kann nicht durch das Standardmodell erklärt werden und wäre ein Hinweis für Neue Physik.
Eine untere Grenze für Δm_s wurde bisher von verschiedenen Experimenten bestimmt. Die Kombination der Messungen ergibt eine Ausschlussgrenze von $\Delta m_s > 14.4$ ps^{-1} (mit 95% C.L.). Fehlende Statistik war bisher der limitierende Faktor für alle Experimente, die eine B_s^0-Oszillations-Analyse durchgeführt haben. Bis zum Start des Large Hadron Colliders (LHC, Cern) Ende 2007 sind CDF[1] und DØ die einzigen Experimente, welche die Frequenz der B_s^0-Oszillation messen können. Beide Experimente liegen am Tevatron, einem p-$\bar{p}$ Beschleuniger am Fermilab in Batavia (Illinois, USA).
CDF hat in Run I (1985-1996) eine untere Grenze für Δm_s ermittelt. Der Betrieb vom Tevatron wurde 2001 nach einer Verbesserung des Beschleunigers und der Detektoren wieder aufgenommen (Run II). CDF und DØ haben bei den Winterkonferenzen 2005 die Resultate der ersten B_s^0-Mischungsanalyse mit Run II Daten gezeigt. Die am Tevatron bestimmte untere Ausschlussgrenze für Δm_s liegt derzeit unter dem Welt-

[1] Collider Detector at Fermilab

mittelwert. CDF und DØ erwarten aber eine höhere Sensitivität bis zum Sommer aufgrund mehrerer Verbesserungen der Analyse. Eine der Verbesserungen der CDF Analyse wird in dieser Doktorarbeit beschrieben.
Diese Arbeit befasst sich mit der Entwicklung eines Algorithmus, der b- und $\bar{b}$-Quarks unterscheiden kann (*b-Flavour Tagger*). Dazu wird mit Hilfe eines neuronalen Netzes eine Größe entwickelt, die anhand von Eigenschaften einer rekonstruierten Teilchenspur die Wahrscheinlichkeit dafür angibt, dass dieses Teilchen aus einem B-Hadron-Zerfall stammt (B-Spurwahrscheinlichkeit). Diese Wahrscheinlichkeit wird dann in einem weiteren neuronalen Netz dazu benutzt, um die Wahrscheinlichkeit zu bestimmen, dass ein Jet die Zerfallsprodukte eines B-Hadrons enthält (B-Jets). Der Algorithmus des auf neuronalen Netzen basierenden Jet-Charge-Taggers (neuronaler Jet-Charge-Tagger) wählt aus einem Ereignis den Jet mit der höchste Wahrscheinlichkeit aus. Er berechnet die Jet-Ladung aus den Ladungen der Spuren im Jet gewichtet mit der jeweiligen B-Spurwahrscheinlichkeit. Die Leistungsfähigkeit des Taggers wurde auf Daten gemessen und ist um 30% besser als die eines vergleichbaren Taggers, der Schnitte zur Selektion der Tagging-Jets benutzt.
Für die Entwicklung eines neuronalen Netzes ist es wichtig, dass die simulierten $b\bar{b}$-Ereignisse die Daten genau beschreiben. Zu diesem Zweck wurden die verschiedenen $b\bar{b}$ Produktionsmechanismen in Monte-Carlo-Daten untersucht, und eine verbesserte Simulation eingeführt, die die experimentellen Daten besser beschreibt.

Die B_s^0-Mischungsanalyse

Um die B_s^0-Oszillationsfrequenz zu bestimmen, muss die zeitabhängige Asymmetrie zwischen der Anzahl von gemischten und ungemischten B_s^0-Kandidaten gemessen werden. Diese Analyse benötigt die Rekonstruktion des Signals in einem Flavour-Eigenzustand, die Messung der Zerfallszeit und die Identifikation der Ladung des b-Quarks, das das B_s^0-Meson enthielt als es erzeugt wurde.
Die Signifikanz für ein B_s^0-Oszillationssignal ist nährungsweise

$$Sig = \sqrt{\frac{S\epsilon D^2}{2}}\sqrt{\frac{S}{S+B}}\; e^{-\frac{(\Delta m_s \sigma_t)^2}{2}} \tag{1}$$

wobei S die Anzahl der rekonstruierten B_s^0-Kandidaten, ϵD^2 die Leistungsfähigkeit des Taggers, B die Anzahl der Untergrundereignisse und σ_t die Auflösung der Messung der Eigenzeit des B_s^0-Mesons ist. Die Eigenzeitauflösung hängt von der Zerfallslängen- und der Impulsauflösung ab. Die Signifikanz nimmt mit wachsendem Δm_s ab. Um schnelle Oszillationen aufzulösen, benötigt man eine exzellente Zerfallslängen- und Impulsauflösung, niedrige Fehltag-Raten und hohe B_s^0-Rekonstruktionsreinheit.
Am Tevatron werden b-Quarks meist paarweise erzeugt und fragmentieren unabhängig

voneinander in B-Hadronen. Das B_s^0-Meson wird durch die Kombination seiner Zerfallsprodukte zu einem Vertex rekonstruiert. Die Eigenzeit wird aus dem Transversalimpuls und der Zerfallslänge in der Transversalebene berechnet. Die Teilchenspuren in der Umgebung des B_s^0-Kandidaten definieren die *Same-Side* des Ereignisses. Die übrigen Teilchenspuren gehören zur *Opposite-Side* und werden benutzt, um das zweite B-Hadron des Ereignisses inklusiv zu rekonstruieren. Die Ladung des B_s^0-Kandidaten beim Zerfall wird aus den Zerfallsprodukten bestimmt. Die Ladung bei der Produktion kann bestimmt werden durch Fragmentationspuren, die den Übergang $\bar{b} \to B_s^0$ begleiten, und Zerfallsprodukte und Fragmentationspuren, die sich auf der Opposite-Side befinden.
Der Jet-Charge-Tagger ist ein Opposite-Side-Tagger, der die Ladungskorrelation zwischen dem b-Quark und den Spuren in seiner Umgebung ausnutzt. Er zeichnet sich durch eine große Effizienz aus. Eine Verbesserung der Reinheit führt deshalb zu einem signifikanten Gewinn der Leistungsfähigkeit.

Der Monte-Carlo-Datensatz

Die $b\bar{b}$-Ereignissimulation, die bisher für Tagginguntersuchungen benutzt wurde, enthält nur Prozesse, in denen zwei schwere Quarks an der harten Streuung teilnehmen (*Flavour-Creation*, Abb. 1a und Abb. 1b). Es wurde festgestellt, dass diese Simulation nicht ausreicht, um $b\bar{b}$-Ereignisse in Daten zu beschreiben. Ein Monte-Carlo-Datensatz, der Flavour-Erzeugung und zusätzliche Prozesse enthält, wird in dieser Arbeit eingeführt. Die zusätzlichen Prozesse im Rahmen des Parton-Schauer-Modells sind *Flavour-Excitation* (FE), bei dem nur ein b-Quark an der harten Streuung teilnimmt (Abb. 1d), und *Gluon-Splitting* (GS), bei dem kein schweres Quark an der harten Streuung teilnimmt (Abb. 1e und 1f). Die drei verschiedenen Prozesse entsprechen verschiedenen Korrelationen zwischen den schweren Quarks.
Der neuer Monte-Carlo-Datensatz beschreibt die Daten besser als die Simulation, die nur Flavour-Creation-Prozesse enthält (Abb. 1, rechts). Außerdem unterscheidet sich in den zwei Sätzen die Schätzung der Leistungsfähigkeit des Opposite-Side-Tagger. Die Simulation, die auschießlich Flavour-Creation-Prozesse enthält, ergibt einen zu optimistischen Wert.
Die verbesserte Simulation sagt voraus, dass in etwa 40% der Ereignisse, die durch ein Lepton und eine verschobene Spur getriggert werden (ℓ+SVT Sample), nur ein b-Quark in dem Akzeptanzbereich des Detektors zu finden ist. Die Trennung von B-Hadron-Spuren aus Fragmentationsspuren ist nur schwer durch Schnitte möglich. Mit Hilfe der in dieser Arbeit eingeführten B-Spurwahrscheinlichkeit kann die Trennung besser durchgeführt werden.

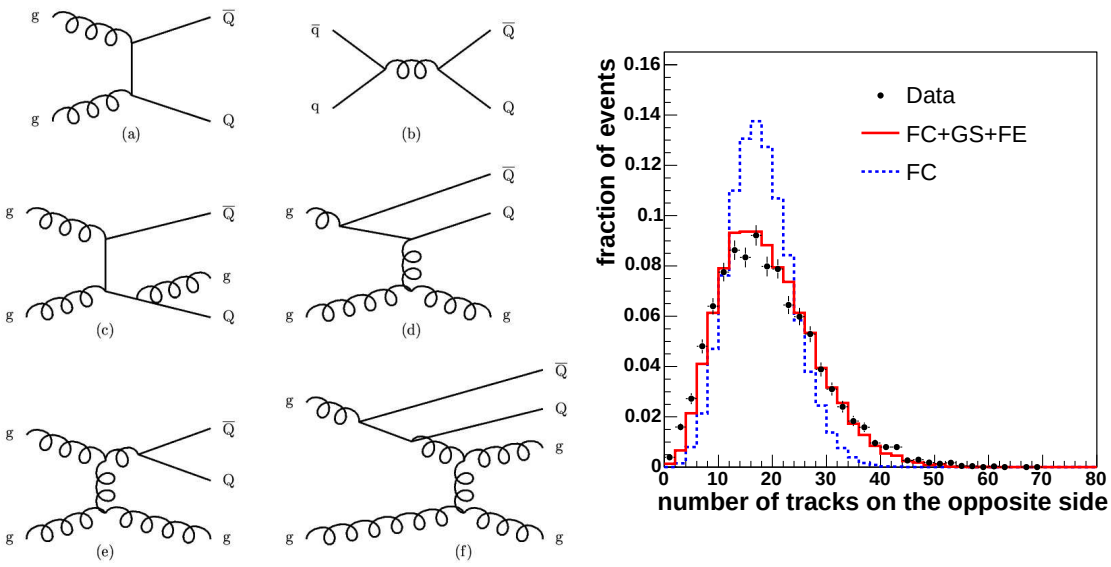

Abbildung 1: *Links:* Feynmandiagramme der wichtigsten $b\bar{b}$-Produktionsprozesse, *(a)*, *(b)* und *(c)* sind Flavour-Creation-Prozesse, *(d)* ist ein Flavour-Excitation-Diagramm, *(e)* und *(f)* entsprechen Gluon-Splitting-Prozessen. *Rechts:* Vergleich von Daten und zwei Monte-Carlo-Datensatz für die Verteilung der Anzahl von Spuren, die auf der Opposite-Side zu finden sind. Der Datensatz mit zusätzlichen Prozessen beschreibt die Daten erheblich besser.

B-Spur- und B-Jet-Wahrscheinlichkeit

Anhand des Monte-Carlo-Datensatz wurden die B-Spur- und die B-Jet-Wahrscheinlichkeit entwickelt. Die B-Spurwahrscheinlichkeit wird durch die Kombination von Informationen, die spezifisch für B-Hadron-Zerfallsprodukte sind, mit Hilfe eines neuronalen Netzes gebildet. Zu diesen Informationen gehören unter anderem der Abstand oder die Richtung der Spur bezüglich der Jet-Achse. Die Wahrscheinlichkeitsvariable trennt erfolgreich Teilchenspuren, die aus einem B-Hadron-Zerfall stammen, von dem Untergrund sowohl in simulierten Ereignissen als auch in experimentellen Daten.
Aufbauend auf der B-Spurwahrscheinlichkeit werden mehrere Jet-Variablen definiert, z.B. die Spur mit der höchsten B-Spurwahrscheinlichkeit im Jet oder die Summe der Wahrscheinlichkeiten aller Spuren. Diese werden wiederum mit Hilfe eines neuronalen Netzes kombiniert und ergeben eine B-Jet-Wahrscheinlichkeit. Diese kann Jets, die aus einem b-Quark stammen, vom Untergrund sowohl in Daten als auch im Monte-Carlo (Abb. 2, links) trennen. Die Separationsleistung der B-Jet-Wahrscheinlichkeitsvariablen ist besser als die von einer zuvor benutzen Variablen, die nur auf dem Abstand der Spuren zum Primärvertex basiert (Abb. 2, rechts).
Es werden drei Jet-Typen definiert: Jets mit einem rekonstruierten Sekundärvertex, Jets ohne rekonstruierten Sekundärvertex und mit mindestens einer Spur, die ei-

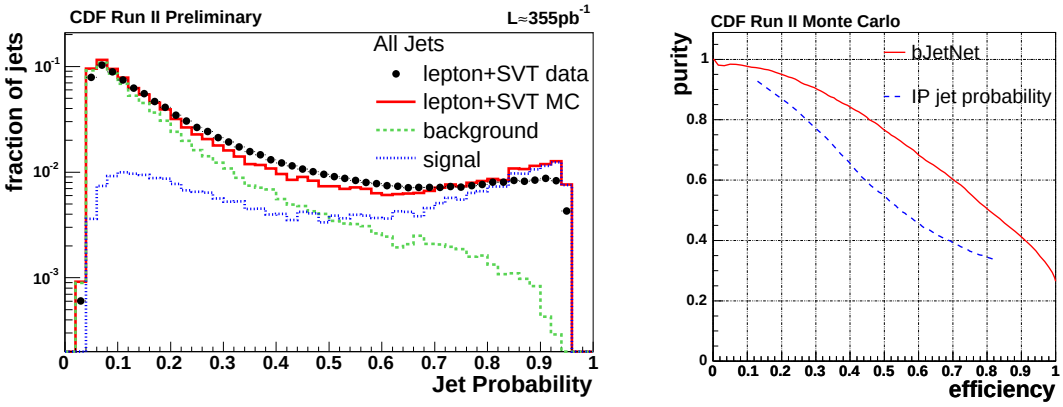

Abbildung 2: Verteilung der B-Jet-Wahrscheinlichkeit für alle rekonstruierten Jets (*links*) und Vergleich der Separationseistung der B-Jet-Wahrscheinlichkeitsvariablen mit einer Variable, die nur Informationen über den Abstand der Spur zum Primärvertex verwendet (*rechts*).

ne B-Spurwarscheinlichkeit größer als 50% besitzt; Jets ohne rekonstruierten Sekundärvertex und ohne eine Spur mit einer Wahrscheinlichkeit größer als 50%. Als Tagging-Jet für den neuronalen Jet-Charge-Tagger wird der Jet im Ereignis mit der höchsten B-Jet-Wahrscheinlichkeit ausgewählt.

Neuronaler Jet-Charge-Tagger

Die Effizienz ϵ eines Taggers gibt an, für welchen Anteil der Ereignisse eine Aussage über die Produktionsladung des b-Quarks gemacht werden kann. Die *Dilution* D misst, wie wahscheinlich es ist, dem Ereignis die richtige Ladung zuzuordnen. Die Taggingleistung ist durch $\epsilon \cdot D^2$ gegeben.
Der neuronale Jet-Charge-Tagger verwendet die B-Jet-Wahrscheinlichkeit P_{nn}, um einen B-Jet zu selektieren, und nutzt die B-Spurwahrscheinlichkeit als Gewicht für die Berechnung der Jetladung Q_{jet}. Die Leistungfähigkeit des Taggers wird mit dem ℓ+SVT Datensatz gemessen. Um eine größere Leistung zu erreichen, wird der Datensatz in Bereiche unterteilt, die unterschiedlichen Werte von $|Q_{jet}| \cdot P_{nn}$ und unterschiedlichen Jet-Typen entsprechen. Die Leistungsfähigkeit wird in jedem Bereich getrennt gemessen, und die Summe ergibt die kombinierte Leistungsfähigkeit, die größer ist, als die auf dem ungeteilten Datensatz.
Das Resultat für die kombinierte Leistungsfähigkeit ist

$$(0.917 \pm 0.031)\% \quad \text{für das } e\text{+SVT Datensatz}$$
$$(0.938 \pm 0.029)\% \quad \text{für das } \mu\text{+SVT Datensatz}$$

Die Leistung des neuronalen Jet-Charge-Taggers ist um 30% besser als die des schnittbasierten Taggers, der bei der B_s^0-Oszillations-Analyse von CDF bisher benutzt wurde. Die Dilution des Taggers wurde als Funktion von $|Q_{jet}| \cdot P_{nn}$ und des Jet-Types parametriesiert, um die Bestimmung der Dilution für jedes Ereignis zu ermöglichen. Dadurch werden Ereignisse, die eine potentiell höhere Dilution besitzen, stärker gewichtet. Da so die Ereignisse, die mehr Information besitzen stärker berücksichtigt werden, erhöht sich die Sensitivität der B_s^0-Oszillationsanalyse. Der in dieser Arbeit vorgestellte Jet-Charge-Tagger wird zur Bestimmung der unteren Grenze der B_s^0-Mischungsfrequenz der Analyse in diesem Sommer benutzt werden.

A Neural Jet Charge Tagger for the Measurement of the B_s^0-$\bar{B}_s^0$ Oscillation Frequency at CDF

Claudia Lecci

Ph.D. Thesis
Faculty for Physics
University of Karlsruhe (TH)

Supervisor: Prof. Dr. M. Feindt, Institut für Experimentelle Kernphysik, University of Karlsruhe (TH)

Contents

1 Introduction **5**

2 B_s^0 Oscillation – Theory Overview **9**

2.1 The CKM Matrix **9**

2.2 The Unitarity Triangle **10**

2.3 B_s^0 Mixing **11**

2.4 Status of B_s^0 Mixing Measurements **13**

3 The CDF Experiment **15**

3.1 Fermilab Accelerator Complex **15**

3.2 The CDF detector **17**

3.2.1 Tracking System 18

3.2.2 Calorimeters 19

3.2.3 Muon System 19

3.3 Trigger System and DAQ **20**

3.3.1 Two Track Trigger 21

3.3.2 Lepton and Displaced Track Trigger 21

3.4 Event Reconstruction **22**

3.4.1 Tracking 22

3.4.2 Primary Vertex 23

3.4.3 Reconstruction and Tagging of Jets 23

4 Outline of B_s^0 Oscillation Analysis **27**

4.1 Signal Reconstruction **30**

4.1.1 Semileptonic Channel 30

4.1.2 Hadronic Channel 31

4.2 Proper Time Reconstruction **31**

4.3 b-Flavour Tagging **33**

4.3.1 Tagging Power ϵD^2 35

4.3.2 Same Side Tagging 36
4.3.3 Opposite Side Tagging 37
4.3.4 Remarks with Respect to B-Factories 39
4.4 Extraction of Δm_s 40

5 Understanding $b\bar{b}$ Events with Simulation 45

5.1 $b\bar{b}$ Pair Production at Tevatron 45
5.2 Samples Description 47
5.2.1 B Candidate Reconstruction 47
5.3 Comparison of Monte Carlo Samples 49
5.3.1 Quark Correlation 49
5.3.2 Comparison to Data 52
5.4 Detector Description in Simulated Events 55
5.5 Event Shape Study 57
5.5.1 Same Side 57
5.5.2 Opposite Side 61
5.6 Jet Clustering Algorithm Evaluation 63
5.6.1 Cone Clustering 64
5.6.2 Mass Clustering 65
5.7 Summary 68

6 b-jet Neural Network 71

6.1 Sample Description 71
6.1.1 Event Pre-Selection 72
6.1.2 Same Side B-Daughters Removal 73
6.1.3 Jet Reconstruction 75
6.2 Neural Network based Track Probability 76
6.2.1 Choice of Input Variables 77
6.2.2 Performance 82
6.3 b-jet Selection Neural Network Variable 82
6.3.1 Choice of Input Variables 84
6.3.2 Performance 90
6.4 b-Jet Probability with Mass Clustering 92
6.5 Summary 95

7 Neural Network Based Jet Charge Tagger 99
7.1 Tagger Setup 99
7.2 Q_{jet} Computation 100
7.3 Measurement of the Tagging Power $\mathcal{T}$ 100
7.3.1 D_{raw} and D_{true} 102
7.3.2 Binned Dilution 102
7.3.3 Effective Dilution 106
7.3.4 Results 107
7.4 Exclusion of Events Tagged by the Soft Lepton Taggers 108
7.5 Further Studies 111
7.5.1 $D(B^0)$ and $D(B^+)$ 111
7.5.2 Alternative Jet Charge Tagger Setup 112
7.5.3 Jet Charge Tagger with Mass Clustering 112
7.6 Summary 114
8 Conclusion and Outlook 115
A True Asymmetry and Measured Asymmetry 117
B Production of the Monte Carlo Samples 119
B.1 Flavour Creation Sample 119
B.2 All $b\bar{b}$ Production Processes Sample 119
C Optimisation of the Track Probability 121
C.1 Network Training 122
C.2 Separate Training for COT only Tracks 125
D Optimisation of the Jet Probability 127
E B-jet Selection Likelihood Variable 131
E.1 Choice of Input Variables 131
E.2 Likelihood Ratio 138
E.3 Performance 139
F Likelihood based Jet Charge Tagger 141
G Further Development of the Jet Charge Tagger 145
G.1 Track Probability Optimisation 145

G.2 Tagger Performance **149**

H Tagger Usage Instructions **151**

H.1 The TrackProbNN Class **151**
H.1.1 How to Use TrackProbNN in AC++ 152
H.2 The NNJetQTagger Class **153**
H.2.1 How to Use NNJetQTagger in AC++ 154
H.3 How to Compile BottomTaggers **156**
H.3.1 Caveat 158

Bibliography **159**

Acknowledgements **165**

Chapter 1

Introduction

The transition of neutral mesons between particle and anti-particle is called *oscillation* (or *mixing*). This phenomenon is explained in the Standard Model of particle physics as a consequence of the fact that the eigenstates of the weak interaction are not mass eigenstates.
Mixing was predicted for the K^0-$\bar{K}^0$ system in 1955 and observed in 1956. The mixing of neutral b-mesons has been observed for the first time in 1987 for the B_d^0 system, which is composed of a $\bar{b}$- and d-quark. The frequency of the B_d^0 meson oscillation, Δm_d, has been measured rather precisely by several experiments and the world average currently is $\Delta m_d = 0.502 \pm 0.007$ ps^{-1} [1].
The oscillation in the B_s^0 system, composed of $\bar{b}s$, has been observed, but its frequency Δm_s is one of the few parameters of the Standard Model which have not yet been measured. It is predicted that the B_s^0 oscillations are very fast, the value of Δm_s being of the order of 10 ps^{-1}. A measured value larger than 25 ps^{-1} cannot be explained by the Standard Model and would therefore be an exciting indication of the existence of New Physics.
Lower limits on Δm_s have been provided so far by different experiments, resulting in a combined limit of ($\Delta m_s > 14.4$ ps^{-1} at 95% C.L.) [1]. The statistics was the limiting factor so far for all the experiments that performed a B_s-mixing analysis. Until the start of the operation of the Large Hadron Collider (LHC) at the end of 2007, the only experiments able to measure the frequency of the B_s^0 oscillation are CDF[1] and DØ, located a the Tevatron Collider for protons and anti-protons at Fermilab, in Batavia (Illinois, USA).
A limit on Δm_s was set by CDF in the first phase of the Tevatron operation (Run I). The Tevatron was shut down during an upgrade period which lasted 5 years. The operation restarted at the end of 2001 (Run II). CDF and DØ have shown at the Winter Conferences 2005 the results of the first search for B_s^0 oscillation with Run II data

[1]Collider Detector at Fermilab

[2, 3]. The accessible range of values of Δm_s is below the world average as of now, but both CDF and DØ expect a better sensitivity by the summer because of several analysis improvements. One of the improvements for the CDF analysis is offered by the work presented in this thesis.
An indirect measurement of Δm_s is given by the measurement of the width difference between the heavy and light B_s eigenstates, $\Delta\Gamma_s$. This quantity has been measured by CDF in 2004 [4] and resulted, within the Standard Model, in $\Delta m_s = 125^{+65}_{-55}\,\mathrm{ps}^{-1}$. The high value for Δm_s could already imply New Physics contributions. Higher statistics is needed to rule out the possibility of a statistical fluctuation.

The measurement of the B_s^0 mixing frequency is a delicate issue which requires high performance from the experiment, very good understanding of $b\bar{b}$ event topology and production mechanisms, good understanding of the detector and precise reconstruction.
In order to measure Δm_s, events in which a $b\bar{b}$ pair was created need to be extracted from the multitude of other events produced in the CDF detector. About 1.5 events in 1000 contain a $b\bar{b}$ pair. A smart solution to this problem was found by the CDF experiment with the introduction of a hardware trigger, the Silicon Vertex Tracker (SVT), which exploits the long lifetime of b-hadrons. The products of the decay of b-hadrons are displaced with respect to the primary interaction point. The displacement is measured at trigger level with high precision, thus allowing a clean selection of $b\bar{b}$ events. The SVT trigger is unique in a hadronic environment.
The B_s^0 meson is reconstructed by finding its decay products among the many particles produced in the $p\bar{p}$ collision. The distance that the B_s^0 travels before decaying is measured and the flavour of the B_s^0 is tagged at production and at decay time, i.e. it is decided if the meson contained a b- or a $\bar{b}$-quark when it was produced and when it decayed. The experimental techniques which identify the b-quark flavour are called *b-flavour tagging*. In particular the methods that tag the flavour of the B_s^0 by exploiting the information relative to the other b-quark in the event are called *opposite side taggers*.
The identification of the other b-hadron in the event presents several difficulties. Often this is produced in the forward direction and its decay products cannot be detected. A simulated $b\bar{b}$ event in the CDF detector is shown in fig. 1.1. The final state particles originating from the decay of the $\bar{B}_s^0$ and B_d^0 mesons are indicated. The $\bar{B}_s^0$ is fully reconstructed, while not all the B_d^0 decay products are detected. The two b-hadrons do not form back-to-back jets, on the contrary their decay products are quite spread in the detector. This renders the identification of tracks from the opposite side b-hadron complicated.
The identification at high efficiency of the decay products of the opposite side b-hadron and their distinction from all the other tracks is a complex task. The purpose of the

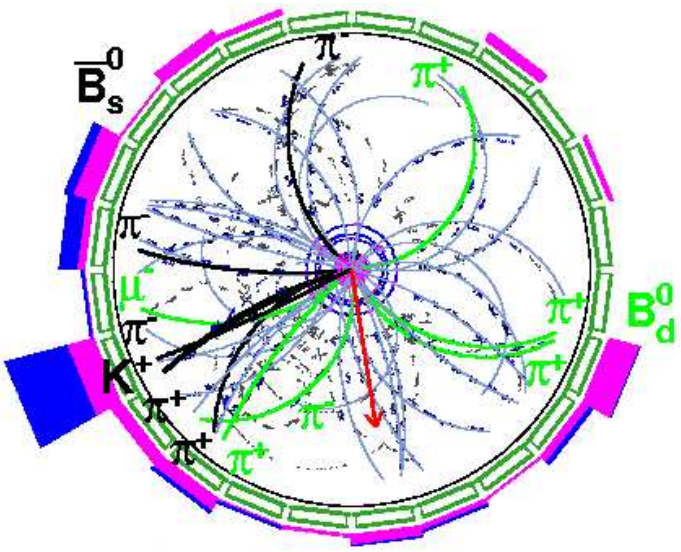

Figure 1.1: Display of a simulated $b\bar{b}$ event in the r-ϕ view of the CDF detector. The arrow indicates the direction of the missing momentum. In this example the b-quark forms a $\bar{B}_s^0$ meson, which decays via the following chain: $\bar{B}_s^0 \rightarrow D_s^+\pi^-$, $D_s^+ \rightarrow K_S^0 K^+\pi^+\pi^-$, $K_S^0 \rightarrow \pi^+\pi^-$. Its final state particle are indicated with bold black font. The opposite side $\bar{b}$-quark forms a B_d^0 meson. Its final decay products (bold green) are only partially reconstructed in the detector. The other tracks (this blue lines) correspond to particles from the fragmentation and the underlying event.

work presented in this thesis is to solve this problem and to apply the result to flavour tagging. In order to achieve a high efficiency, as much information as possible on the decay products of the b-hadron has been conveyed. The optimal combination of variables has been provided by a Neural Network and the result yielded a powerful variable which gives a probability that a track originated from the b-hadron decay. The b-track probability has been employed to define jet-level variables, which have also been combined with a Neural Network to construct a b-jet probability variable. A similar approach has been adopted in the DELPHI experiment[2] by means of the BSAURUS package based on Neural Networks [5]. This is a comprehensive tagging package that combines all possible information at the track level to provide a probability for the flavour of the B_s^0 candidate. BSAURUS has been conceived in Karlsruhe, where a huge expertise on Neural Networks has consolidated over the past 10 years. The Karlsruhe CDF B group is now in the process of transferring this technology to CDF. The b-track and b-jet probability variables designed for CDF are the first

[2]DELPHI was one of the four experiments operating at LEP (Large Electron-Positron collider), at CERN (Geneva, Switzerland) between 1989 and 2000.

official Neural Network tools adopted by the CDF B group and constitute the first step toward a combined Neural Network based tagging package for an experiment at a hadron collider. The differences between the clean e^+e^- collider environment and the crowded $p\bar{p}$ collider environment make it very challenging for CDF to reach a similar performance. Nevertheless the application of Neural Networks to CDF events proves successful, as it is going to be demonstrated in this thesis.
The probability variables have been employed in an opposite side tagger, the *Jet Charge Tagger*, to select the jet that is most likely to contain the decay products of the opposite side b-hadron and to extract the b-quark flavour from the weighted sum of the charge of the tracks in the jet. The new tagger proved to be more powerful than the tagger currently in use at CDF and it is going to improve the CDF B_s^0 mixing analysis, which will be updated for Summer 2005.

In Chapter 2 a brief theoretical overview on B_s^0 mixing is given. Its explanation in the Standard Model in the framework of the CKM quark-mixing matrix is provided. The measurement of Δm_s performed so far are summarised. The experimental setup for the measurement of Δm_s is illustrated in Chapter 3, where the CDF detector is described focusing on the elements relevant for the B_s^0 mixing analysis. Chapter 4 reports about the B_s^0 mixing analysis presented by CDF to the Winter Conferences 2005 and it explains the fundamental elements: signal reconstruction, proper time measurement and flavour tagging.
Intense simulation studies have been carried on to understand the features of the opposite side of $b\bar{b}$ events, with the aim of improving the existing cut based Jet Charge Tagger. An important contribution brought by these studies to the CDF B group was to point out that the simulation with the $b\bar{b}$ production processes traditionally in use is not sufficient to correctly represent the features of the opposite side. Additional processes are needed to properly describe data, as it is illustrated in Chapter 5. The simulation with additional processes has been employed to build the b-track and the b-jet probability variables. These are presented in Chapter 6. The Neural Network based Jet Charge Tagger, which profits from the probability variables, and its tagging power measured on data are the subject of Chapter 7. Finally Chapter 8 draws the conclusions.
The development of the Neural Network based Jet Charge Tagger involved the generation and simulation of Monte Carlo samples, the training of Neural Networks and studies of methods alternative to Neural Networks for the combination of variables. In addition, different setups were considered for the tagger, which did not prove more powerful than the Jet Charge Tagger presented in Chapter 7. Technical details and additional information are found in the Appendices.

Chapter 2

B_s^0 Oscillation – Theory Overview

The measurement of the frequency of B_s^0 oscillation motivates the work presented in this thesis. The mixing of neutral B mesons is a phenomenon explained in the Standard Model of particle physics as a consequence of quark flavour mixing. Experimental results have not yet confirmed or rejected the Standard Model prediction for Δm_s. In this chapter the theoretical background of B_s^0 mixing is outlined.

2.1 The CKM Matrix

In the Standard Model the quark mass eigenstates (d, s, b) are different from the flavour eigenstates (d', s', b'). The two bases are related through the 3×3 complex Cabibbo-Kobayashi-Maskawa matrix V_{CKM}

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

The existence of only three generations of quarks implies the unitarity of the CKM matrix. The unitarity condition

$$V_{CKM}^{\dagger} V_{CKM} = \mathbb{1} = V_{CKM} V_{CKM}^{\dagger} \tag{2.1}$$

imposes relations among the elements, so that all of them can be expressed in terms of three real parameters and one purely imaginary phase. The Wolfenstein parameterisation introduces the four real parameters λ, A, ρ and η:

$$V_{CKM} \simeq \begin{pmatrix} 1 - \lambda^2/2 & \lambda & \lambda^3 A(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & \lambda^2 A \\ \lambda^3 A(1 - \rho - i\eta) & -\lambda^2 A & 1 \end{pmatrix} + O(\lambda^4)$$

The element $|V_{us}|$ is well determined by the measurements of semileptonic kaon decays and is known to an accuracy of better than 1% [1]. Semileptonic b-meson decays to charmed mesons provided the measurement of $|V_{cb}|$ with a precision better than 5% [1]. Thus the Wolfenstein parameters λ and A are rather well determined:

$$\lambda = |V_{us}| = 0.2200 \pm 0.0026 \tag{2.2}$$

$$A = \left|\frac{V_{cb}}{V_{us}^2}\right| = 0.80 \pm 0.04 \tag{2.3}$$

The experimental uncertainties on ρ and η are at a level of 20%. The parameter η, which is associated with the phase, causes differences in the description of processes involving quarks and anti-quarks. In the Standard Model this phase is the only source of CP violation.

2.2 The Unitarity Triangle

The relations among the CKM elements that follow from the unitarity condition (eqn. 2.1) can be visualised by means of triangles in the complex plane. One of the possible triangles is given by the relation

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

This triangle is called the *Unitarity Triangle* and it is preferred over the others, which are very elongated and more difficult to measure. All triangles have the same area, which is a measure of CP violation in the Standard Model. Figure 2.1 shows the Unitarity Triangle in the complex plane. The sides have been scaled by the approximately real number $V_{cd}V_{cb}^*$. The length of the side $|V_{td}V_{tb}^*/(V_{cd}V_{cb}^*)|$ can be determined from the known values of $V_{tb} \simeq 1$, $V_{cd} \simeq -\lambda$ and $V_{cb} \simeq \lambda^2 A$, and by measuring V_{td}. The determination of V_{td} is an interesting test of the Standard Model since V_{td} is sensible to the phase $i\eta$, therefore to CP violation. The role played by the measurement of the B_s^0 mixing in the determination of V_{td} is explained in Section 2.3.

The current status of the experimental determination of the Unitarity Triangle [12] is displayed in fig. 2.2.

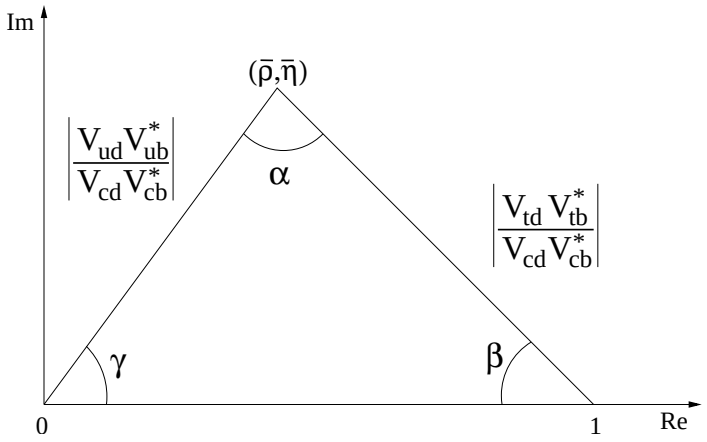

Figure 2.1: The Unitarity Triangle drawn in the complex plane. The quantities $\bar{\rho}$ and $\bar{\eta}$ are equal, respectively, to $\left(1-\lambda^2/2\right)\rho$ and $\left(1-\lambda^2/2\right)\eta$.

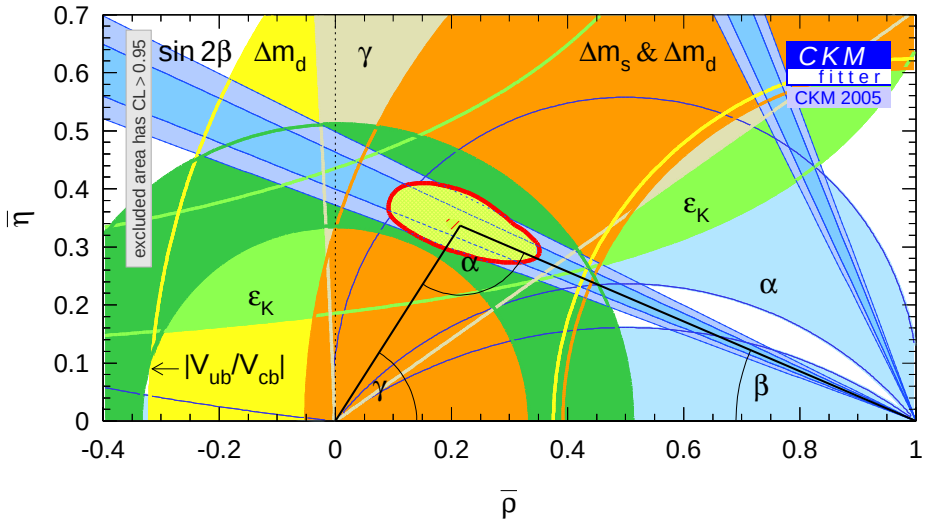

Figure 2.2: Current status of the determination of the Unitarity Triangle [12].

2.3 B_s^0 Mixing

The neutral meson system B^0-$\bar{B}^0$ is composed by the flavour eigenstates

$$|B_q^0\rangle = \bar{b}q \tag{2.5}$$

$$|\bar{B}_q^0\rangle = b\bar{q} \tag{2.6}$$

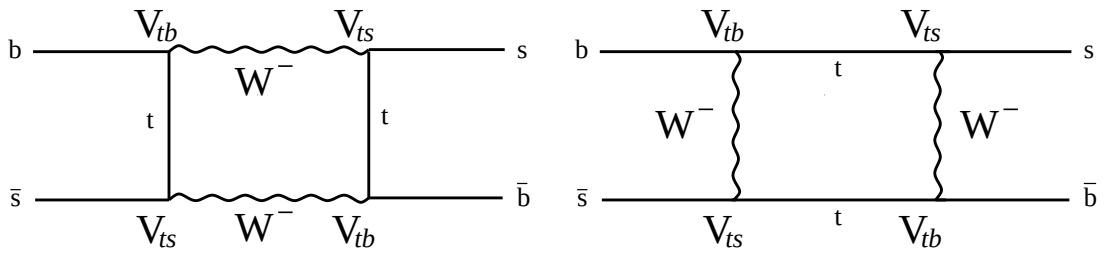

Figure 2.3: Feynman diagrams describing the B_s^0 mixing at the lowest order in the Standard Model. The exchange of a *c*- or a *u*-quark instead of a *t*-quark is suppressed. B_d^0 mixing happens via the same processes involving a *d*-quark instead of a *s*-quark.

where q can be either a *d*- or an *s*-quark. The flavour eigenstates are superpositions of heavy and light mass eigenstates B_H and B_L:

$$|B_q^0\rangle = \frac{1}{\sqrt{2}}\left(|B_{q,H}\rangle + |B_{q,L}\rangle\right) \tag{2.7}$$

$$|\bar{B}_q^0\rangle = \frac{1}{\sqrt{2}}\left(|B_{q,H}\rangle - |B_{q,L}\rangle\right) \tag{2.8}$$

Due to the difference in mass[1], B_H and B_L evolve differently as a function of time, resulting in a time-dependent B^0-$\bar{B}^0$ oscillation with frequency equal to the mass difference $\Delta m_q = m_{q,H} - m_{q,L}$. As a consequence, an initially pure $|B_q^0\rangle$ state may be found at a later time to decay as a $|B_q^0\rangle$, in which case it is *unmixed*, or as a $|\bar{B}_q^0\rangle$, that is *mixed*. The corresponding probability density is

$$P(B_q^0 \to B_q^0, t) = \frac{\Gamma_q}{2} e^{-\Gamma_q t}\left(1 + \cos\Delta m_q t\right) \tag{2.9}$$

$$P(B_q^0 \to \bar{B}_q^0, t) = \frac{\Gamma_q}{2} e^{-\Gamma_q t}\left(1 - \cos\Delta m_q t\right) \tag{2.10}$$

where $\Gamma_q = 1/\tau_q$ and τ_q is the lifetime of the B_q^0 meson.
The oscillation frequency Δm_q can be computed at the lowest order with the box diagram in fig. 2.3. The result of the calculation [7] is

$$\Delta m_q = \frac{G_F^2}{6\pi^2} m_{B_q} m_t^2 \, F\left(\frac{m_t^2}{m_W^2}\right) f_{B_q}^2 B_{B_q} \eta_{QCD} |V_{tb}^* \, V_{tq}|^2 \tag{2.11}$$

where G_F is the Fermi constant, m_{B_q} is the B_q^0 meson mass, m_t is the top quark mass, m_W is the W boson mass, F is the Inami-Lim function which gives the electroweak loop contribution of the top quark without QCD corrections [8], and η_{QCD}

[1]The Standard Model prediction for the decay width difference $\Delta\Gamma_q$ is very small and it allows us to take $\Gamma_{q,L} \simeq \Gamma_{q,H} \simeq \Gamma_q$

is a factor that takes into account QCD corrections to the box diagram [9]. The parameter B_{B_q} and the decay constant f_{B_q} parameterise hadronic matrix elements [10]. A measurement of the B_s^0 (B_d^0) oscillation allows to determine the CKM matrix element V_{ts} (V_{td}). The fairly precise measured value of the B_d^0 oscillation frequency $\Delta m_d = 0.502 \pm 0.007$ ps^{-1} [1] is not enough for the extraction of V_{td} with eqn. 2.11 because of large theoretical uncertainties of the product $f_{B_q}\sqrt{B_{B_q}}$, computed with Lattice QCD techniques. Most theoretical uncertainties cancel out in the ratio between B_s^0 and B_d^0 oscillation frequencies [11]

$$\frac{\Delta m_s}{\Delta m_d} = \frac{m_{B_s} f_{B_s}^2 B_{B_s}}{m_{B_d} f_{B_d}^2 B_{B_d}} \left|\frac{V_{ts}}{V_{td}}\right|^2 \tag{2.12}$$

$$\frac{|V_{ts}|}{|V_{td}|} = \frac{1.01}{\xi}\sqrt{\frac{\Delta m_s}{\Delta m_d}} \tag{2.13}$$

where $\xi = \frac{f_{B_s}\sqrt{B_{B_s}}}{f_{B_d}\sqrt{B_{B_d}}} \simeq 1.24$ is evaluated with Lattice calculations with an uncertainty below 5% [6]. From eqn. 2.12 and the Wolfenstein parameterisation it is clear that B_s^0 oscillations are much faster than B_d^0 oscillations

$$\frac{\Delta m_s}{\Delta m_d} \simeq \frac{1}{\lambda^2} \simeq 20 \tag{2.14}$$

Resolving these rapid oscillations thus poses a serious experimental challenge, an overview of which is given in Chapter 4. The Standard Model does not favour Δm_s values above 25 ps^{-1} [13]. Thus a high frequency would indicate New Physics contributions, for example new particles in the box diagram in fig. 2.3. In this case the independent determination of the angles and the sides of the Unitarity Triangle would lead to an open triangle.

2.4 Status of B_s^0 Mixing Measurements

The first observation of B^0-$\bar{B}^0$ mixing was published in 1987 by the UA1 and ARGUS collaborations [14]. The two experiments performed a time-integrated analysis and many experiments followed this path since then. Time-dependent analyses offer a better sensitivity to the oscillation frequency. However, while Δm_d has been measured with good precision at LEP, at the B-Factories and at the Tevatron, the B_s^0 oscillation frequency Δm_s has not been measured yet. LEP experiments [15], SLD [16] and CDF Run I [17] have performed B_s^0 mixing analyses that set limits on the value of Δm_s. The lack of statistics did not allow to establish an oscillation signal. The combined result is

$$\Delta m_s > 14.4 \text{ ps}^{-1} \tag{2.15}$$

with 95% confidence level [1]. Before the startup of LHC experiments, which is foreseen in 2007, Δm_s can only be measured by the CDF and DØ collaborations. This is because the B-Factories BaBar and BELLE run at the $\Upsilon(4S)$ production threshold, which is lower than the threshold for the production of $B_s^0\bar{B}_s^0$ pairs. The Tevatron is therefore the only operating collider where B_s^0 mesons are created.
CDF and DØ have shown the first result for B_s^0 oscillation with Run II data [2, 3] at the Winter Conferences 2005. The sensitivity is below the current world average as of now, but CDF expects a better sensitivity by the summer due to several analysis improvements, among them the Jet Charge Tagger presented in this thesis, which is introduced in Section 4.3.3 and described in Chapter 7.

Chapter 3

The CDF Experiment

The CDF experiment is located at the *Tevatron* collider at the Fermi National Laboratories (Fermilab), in Batavia/Illinois (USA), where protons and anti-protons circulate in opposite directions in a ring with a diameter of 2 km. The Tevatron is the accelerator with the highest centre of mass energy currently in operation. The first collisions at a centre of mass energy of 1.8 TeV were initiated in 1985. The data collected until 1996 in the so called *Run I* phase amount to 90 pb^{-1} and delivered, among other interesting results, the first experimental evidence of the top quark [18], followed by the precise determination of its mass [19]. Starting in 1996, the accelerator complex was upgraded to increase the centre of mass energy to 1.96 TeV and the instantaneous luminosity. CDF and DØ, the second Tevatron experiment, were upgraded as well. The *Run II* phase started at the end of 2001 and is scheduled until 2009. During this time 4.4 to 8.5 fb^{-1} of data are expected to be collected. The physics programs of CDF and DØ include Higgs searches, Top-Quark physics, rare processes and the measurement of the frequency of the B_s^0 oscillation. In this chapter the experimental setup of the CDF detector is presented with special focus on the facilities relevant to the Δm_s measurement.

3.1 Fermilab Accelerator Complex

The Fermilab accelerator chain is shown in fig. 3.1. The production of protons starts with negatively charged hydrogen ions that are accelerated to 750 keV in a *Cockroft-Walton* pre-accelerator (not shown in the picture). The ions enter then the linear accelerator (*Linac*) where their energy is increased up to 400 MeV. The electrons are removed from the ions by means of a stripper foil before the injection in the *Booster*, a synchrotron of 150 m in diameter. The protons are accelerated to 8 GeV in the Booster and then transferred to the *Main Injector*, which is a multipurpose ring. Part of the protons are delivered to a fixed target area (*Switchyard*), another part is employed

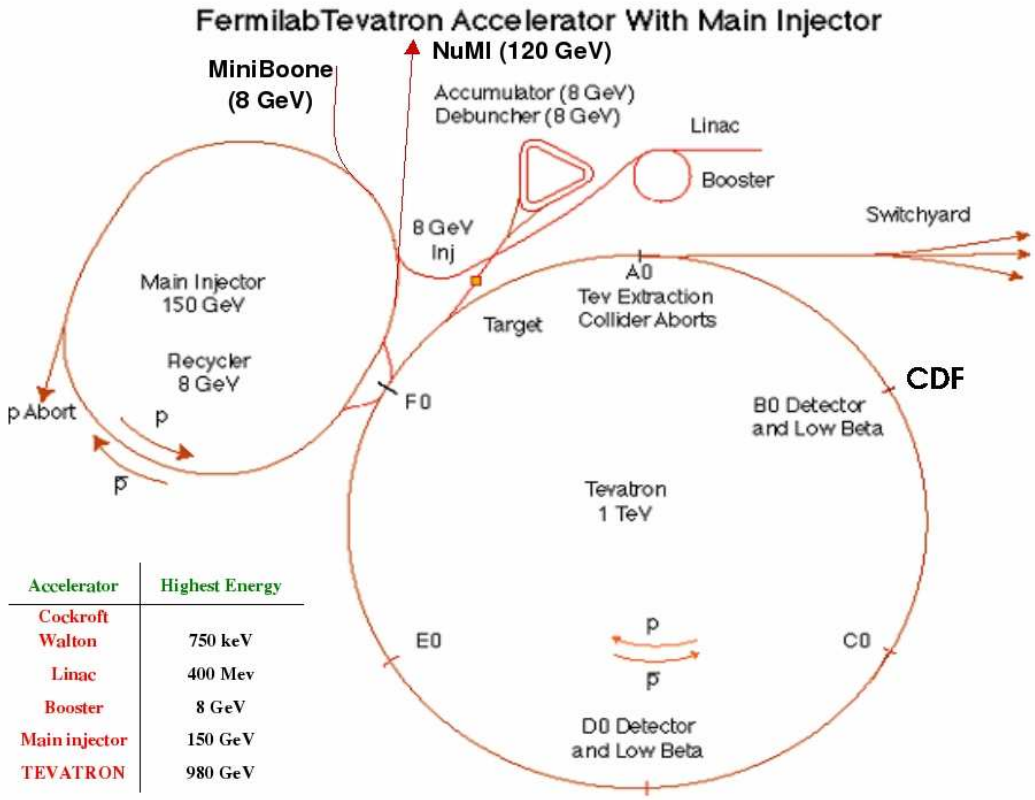

Figure 3.1: Fermilab accelerator chain [20]

for the production of neutrinos by the NuMI and MiniBoone experiments. Some of the protons are accelerated to 120 GeV and sent against a nickel target to produce anti-protons, which are stored in the *Accumulator Ring* before they are injected into the *Recycler*. The Recycler works at the fixed energy of 8 GeV and was designed to re-use the anti-protons left over from a Tevatron store. It serves now as an intermediate storage ring for the anti-protons and allows the build up of the $\bar{p}$ stack while the Tevatron is in collision mode. The Main Injector accelerates the protons and the anti-protons from 8 GeV to 150 GeV and transfers them to the Tevatron, where they are finally brought to an energy of 980 GeV, resulting in a centre of mass energy of $\sqrt{s} = 1.96$ TeV. Protons and anti-protons circulate in the same beam pipe in opposite directions separated by an electrostatic beam. The beams are divided in 3 trains of 12 bunches with a separation of 396 ns and collide in two regions, where the CDF and DØ experiments are located (respectively the BØ and the DØ Tevatron access points).
The integrated luminosity delivered and the amount of data stored by CDF since the start of Run II are shown in fig. 3.2.

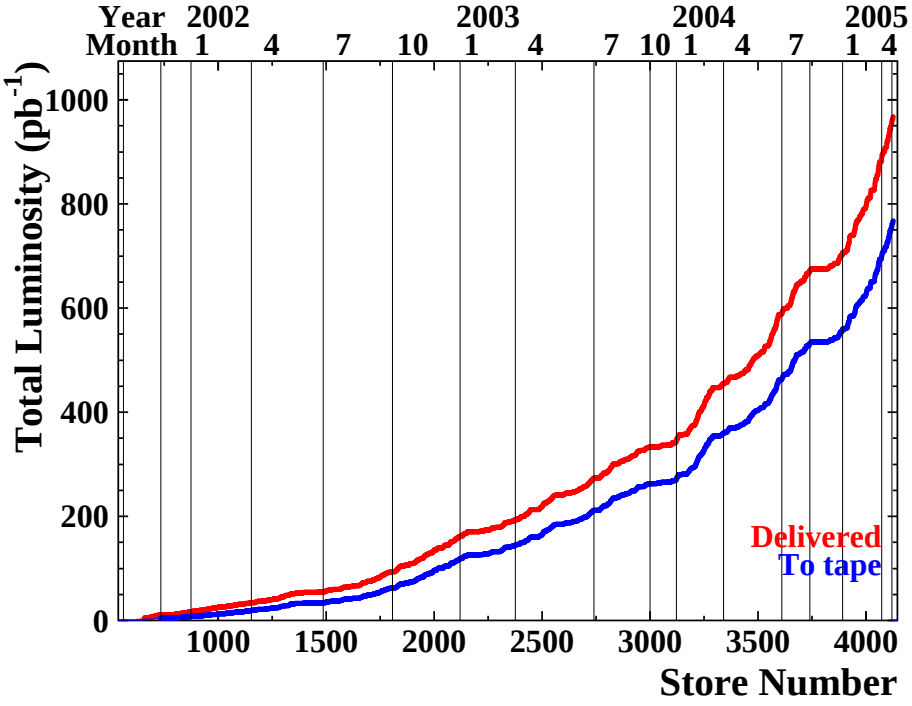

Figure 3.2: Integrated luminosity delivered by the Tevatron (upper line) and stored to tape by CDF (lower line) since the start of Run II (as of 07/05/2005) [21].

3.2 The CDF detector

The CDF Run II detector [23] is a multipurpose detector with cylindrical symmetry with respect to the beam line and forward-backward symmetry with respect to the centre of the detector, i.e. the nominal interaction point. The transverse profile of the interaction region can be approximately described by a circular Gaussian distribution with a typical RMS width of 30 μm. The longitudinal profile is also approximately Gaussian with a typical RMS width of 30 cm. A view of half of the CDF detector is shown in fig. 3.3.

The coordinate system we refer to in the description of the CDF sub-detectors and in the follow up is a cylindrical system. The origin of the system is the centre of the CDF detector. The polar angle θ is measured from the positive z-direction, i.e. the proton direction, pointing eastward from the CDF detector. The azimuthal angle ϕ is measured from the Tevatron plane. It is also convenient to define the *pseudorapidity* $\eta = -\ln(\tan(\theta/2))$, since in hadron-hadron collisions the same number of particles is found in equal intervals of pseudorapidity. As usual in cylindrical coordinates, r measures the distance from the origin measured in the plane perpendicular to the z-direction.

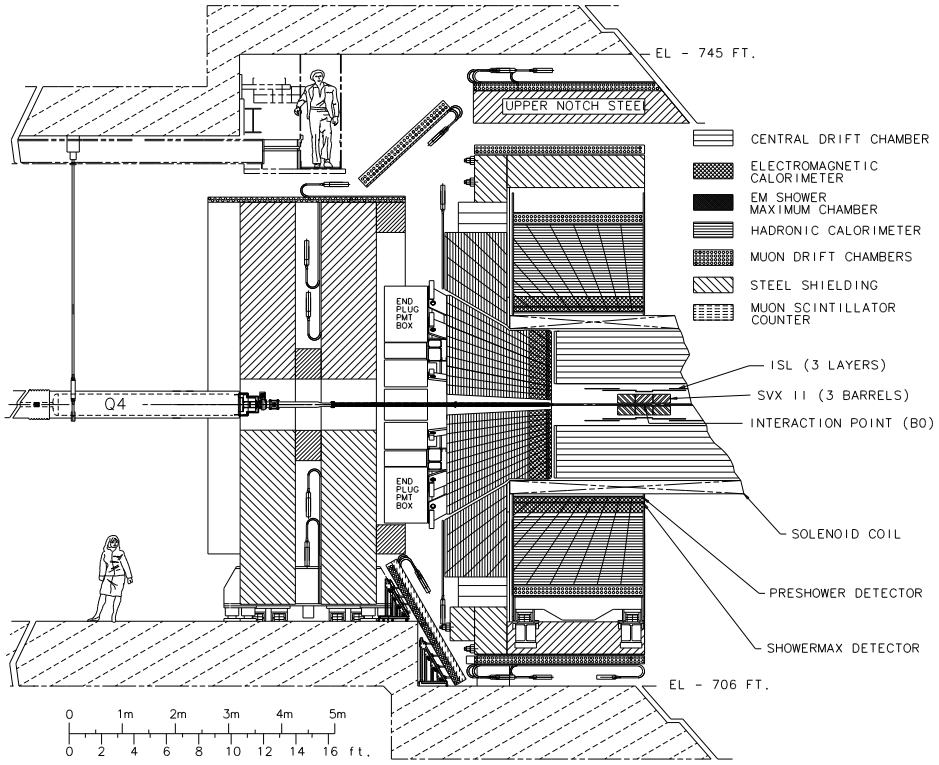

Figure 3.3: Elevation view of one half of the CDF detector [22].

3.2.1 Tracking System

The inner part of the CDF detector is dedicated to the reconstruction of the trajectories of charged particles and the measurement of their momenta. The tracking system is contained in a superconducting solenoid which generates a homogeneous 1.4 T magnetic field parallel to the beam axis.

The silicon micro-strip detector [24] consists of 7-8 layers in a barrel geometry covering the radial region from 1.5 cm to 28 cm. The layer closest to the interaction point is *Layer* 00 (L00), a radiation-hard, single sided detector glued directly onto the beam pipe. The remaining seven layers are double-sided detectors. The five layers after L00 comprise the *Silicon Vertex Detector* of Run II (SVXII) which extends radially from $r_1 = 2.5$ cm to $r_2 = 10.7$ cm. The SVXII measures the track coordinates in r-ϕ and r-z view, thus allowing three-dimensional track reconstruction. The *Intermediate Silicon Layer* (ISL) detector is constituted by one layer in the central region ($|\eta| < 1$) and two layers in the forward region ($1.0 < |\eta| < 2.0$). It plays an important role in linking the hits in the silicon detector with the tracks in the drift chamber.

The *Central Outer Tracker* [25] (COT), is a 3.1 m long open cell drift chamber which covers the radial range from 40 cm to 137 cm. It provides 96 measurement layers,

organised into alternating axial and $\pm 2^\circ$ stereo superlayers. The COT provides coverage for $|\eta| < 1$ and measures the momenta of particles down to 0.4 GeV/c. The energy loss dE/dx of charged particles can be measured as well and is used for particle identification.
A *Time-of-Flight* (TOF) detector [26] is installed in a few centimetres clearance just outside the COT. Its purpose is the measurement of the time-of-flight of particles. The combination of this quantity with the momentum measurement in the COT yields a mass measurement, thus allowing to distinguish among the different particle species. Pions and kaons can be separated with at least one sigma confidence for momenta $p < 1.6$ GeV. This feature is crucial for a clean reconstruction of b-hadron samples and flavour tagging, i.e. in oscillation measurements.

3.2.2 Calorimeters

The solenoid is surrounded by the calorimeters for the detection of showers initiated by photons or electrons (*Electromagnetic Showers*) or by hadrons (*Hadronic Showers*). The calorimeters are segmented along the longitudinal direction to allow the distinction between electromagnetic and hadronic showers. They are sampling detectors, constituted by alternated layers of absorber (lead or iron) and scintillator material. The electromagnetic calorimeter resides in the inner part. It is subdivided into *Central Electromagnetic Calorimeter* [27] (CEM), for the detection of showers in the pseudorapidity range $|\eta| < 1.1$, and *Plug Electromagnetic Calorimeter* [28] (PEM), covering the forward region ($1.1 < |\eta| < 3.64$).
To improve the position and direction measurement of the shower, layers of proportional chambers are located between the solenoid and the CEM (*Pre-Shower Detector*, CPR), for showers initiating in the magnet, and inside the calorimeter at the radial distance where the showers reach the maximum transverse size (*Shower Max Detector*, CES).
The hadronic showers are measured in the outer part of the calorimeter, which is subdivided into a central (CHA) [29], a wall (WHA) and plug (PHA) calorimeter. The first one covers the pseudorapidity region $|\eta| < 0.9$, the second one extends from $|\eta| = 0.7$ to $|\eta| = 1.3$ and the third one detects showers in $1.1 < |\eta| < 3.64$. All particles, except for muons and neutrinos, lose all their energy and are stopped in the calorimeter.

3.2.3 Muon System

The muon system is located outside of the calorimeters and consists of drift chambers and steel absorbers. Since hadrons, photons and electrons are stopped in the calorimeters, muons are the only particles able to reach the muon system. Conse-

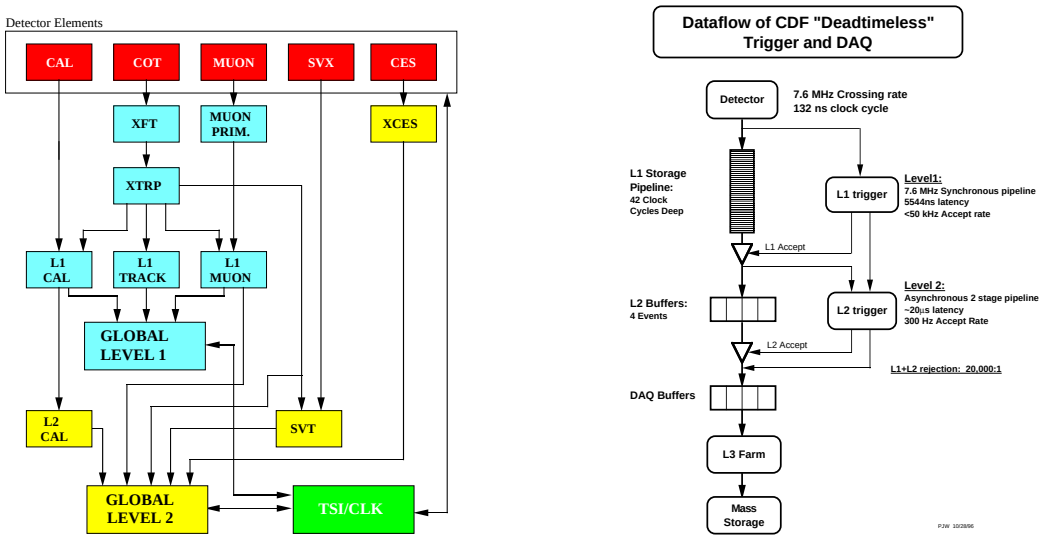

Figure 3.4: *Left:* Scheme of the CDF Trigger System [22]. *Right:* Block diagram of the CDF data flow system [22].

quently these detectors provide a clean identification for muons, which are often produced in the semileptonic decays of b-hadrons. The *Central Muon Detector* (CMU) [30] and *Central Muon Upgrade* (CMP) chambers each provide coverage in the pseudorapidity region $|\eta| < 0.6$. The *Intermediate Muon Detector* (IMU) extends in the region $1.0 < |\eta| < 1.5$. A minimum p_T of 1.4 GeV/c is needed for a muon to reach the muon system.

3.3 Trigger System and DAQ

The trigger and data acquisition systems are designed to accommodate the high rates and large data volume of Run II. Based on preliminary information on tracking, calorimetry and muon system, the output of the first level of the trigger, Level 1, is used to limit the rate for accepted events to a maximum of 50 kHz. The next trigger stage, Level 2, accesses more refined information and additional tracking information from the silicon detector. The accept rate is reduced further to ∼300 Hz. At the last trigger stage, Level 3, the full event reconstruction is performed on a Linux PC farm. This allows the trigger decision to be based on the full event information. In the last stage, the output rate is reduced to ∼75 Hz and the accepted events are written to permanent storage divided in different datasets depending on the met L1/L2/L3 requirements (*Trigger Paths*). A scheme of the detector information accessed by L1 and L2 is shown in fig. 3.4 (left). Figure 3.4 (right) illustrates the event rate reduction

as the events flow from the readout to the storage area.
CDF features triggers designed for B Physics, such as the *Silicon Vertex Tracker* [31] (SVT) at L2. Due to its long lifetime, the b-hadron can travel several hundreds of microns before it decays. The decay vertex is detached from the primary interaction point and its tracks are characterised by a large impact parameter d_0 with respect to the interaction point. The correlation between large track impact parameters and b-hadron lifetime is exploited by the SVT trigger. At L1 the *Extremely Fast Track Finder* (XFT) [32], a trigger with hardware implementation finds tracks in the transverse plane of the COT. Later the SVT, which is also a hardware trigger, matches r-ϕ silicon hits to the XFT tracks, allowing the determination of the track parameters in the r-ϕ plane with a resolution comparable to the offline reconstruction.
Two trigger paths that rely on the SVT decision and are relevant to this thesis are the *Two Track Trigger* (TTT) and the *Lepton and Displaced Track Trigger* (ℓ+SVT).

3.3.1 Two Track Trigger

The TTT requirements yield a sample enriched in $B \to h^+h^-$ decays. The conditions to be satisfied are:

Level 1 at least two XFT tracks with $p_T > 2$ GeV/c, $p_{T,1} + p_{T,2} > 5.5$ GeV/c and with an angle in the transverse plane $\Delta\phi^{1,2} < 135°$

Level 2 at least two SVT tracks with $p_T > 2$ GeV/c and 100 μm $< |\mathrm{d}_0| <$ 1 mm and a good match to the XFT tracks ($\chi^2_{SVT} < 25$)

Level 3 match of the two SVT trigger tracks with those reconstructed in the COT, confirmation of the L2 requirements on p_T and d_0; additional cut $2° < \Delta\phi^{1,2} < 90°$ and $L_{xy} > 200$ μm, where L_{xy} is the decay length in the r-ϕ plane projected onto the direction of the added transverse momenta of the trigger tracks.

3.3.2 Lepton and Displaced Track Trigger

The ℓ+SVT comprises a set of conditions that yield a sample enriched in semileptonic b-hadron decays, i.e. $B \to \ell\nu_\ell h_{SVT}X$, where the lepton ℓ is either a muon or an electron. The requirements to be satisfied are:

Level 1 a muon detected in the CMU and in the CMP, or an electron in the CEM with $E_T > 4$ GeV and ratio of the energies measured in the hadronic calorimeter and in the electromagnetic calorimeter $E_{HAD}/E_{EM} < 0.125$. The lepton is matched to a XFT track with $p_T > 2$ GeV/c

Level 2 at least one SVT track with $p_T > 2$ GeV/c and 120 μm $< |d_0| <$ 1 mm and $\chi^2_{SVT} < 25$. The transverse invariant mass $M_T(\ell, SVT)$, computed using only the sum of the transverse momenta of the lepton and of the SVT track, is smaller than 5 GeV and $5^\circ < \Delta\phi^{1,2} < 90^\circ$. In addition the electron is required to be matched to a shower in the CES with $E_T > 4$ GeV

Level 3 confirmation of the L2 requirements.

3.4 Event Reconstruction

The reconstruction of an event in the CDF detector employs several algorithms and it produces a large number of "objects", such as tracks, calorimeter jets, vertices, TOF-track matches, etc. In this section the reconstruction of the event objects which are relevant for this thesis is briefly described.

3.4.1 Tracking

Track reconstruction begins in the COT. Since the track density is lower far from the interaction point, the pattern recognition first forms line segments from hits in each superlayer. Line segments from the axial layers that are tangent to a common circle are linked together to form a track candidate and the hit positions are fit to a circle. Line segments in stereo layers are then linked to the 2-dimensional track and a helix fit is performed (*COT Standalone Tracking*). A second algorithm based on histogramming exists for the reconstruction of tracks in the COT [33]. The next step is the extrapolation of each COT track into the silicon detector and the addition of hits that are consistent with the track (*Outside-In Tracking*, OI) [34]. A window around the track is established based on the errors on the COT track parameters. If a hit in the outer SVXII layer lies within the window, it is added to the track. A new track fit is then performed, resulting in a new error matrix and a new window. This window is then used to add hits in the next SVXII layer, and the procedure is repeated over all layers. If no hit is found within the search window, the algorithm proceeds to the next layer. An OI track is formed only if at least three r-ϕ hits in the SVXII are associated with the original COT track.
Due to the COT coverage, the OI algorithm reconstructs tracks in the region $|\eta| < 1$. To recover forward tracks, *Silicon Standalone Algorithm* [35] (SISA) uses only hits in the silicon detector. The tracks are formed by combining hits not used by the OI algorithm. In order to reduce the number of hit combinations, a pre-tracking primary vertex finder [36] is run to find the z-coordinate of the primary interaction point. The combinations of hits resulting in tracks incompatible with these z-positions are not pursued by the SISA tracking. This significantly reduces processing time and fake

rate.
In order to recover the COT information for tracks that cross only the first three superlayers, the *Inside-Out Tracking* (IO) has been recently introduced [37]. The SISA tracks define a search road in the COT and the new track is fit constraining the track impact parameter d_0 and the z-coordinate at the point of closest approach to the origin z_0 to those of the SISA track. The silicon hits are then re-fitted using the COT track as seed and the parameters of the new track define the IO track.

3.4.2 Primary Vertex

Many particles in a $b\bar{b}$ event originate from the primary interaction point (*Primary Vertex*, PV). Tracks produced in the decay of long lived particles are displaced with respect to the PV. A good approximation of the PV position is the beam spot. A better determination is provided by the *VxPrim* algorithm [38], an iterative vertex fit. Loosely selected tracks with hits in the COT are combined to form a vertex and after each fit iteration the track with the worst χ^2 is eliminated. The procedure is repeated until no track has a χ^2 larger than a maximum allowed or when the number of tracks in the vertex reaches the minimum allowed value. Although no strict requirement on the track impact parameter is applied, tracks with large displacement are progressively excluded from the vertex fit. Consequently the tracks in the PV in average have a small d_0.
In the follow up the beam spot is mostly used as the primary interaction point, except for applications in which the result of the iterative fit offers valuable information on the tracks (see Section 6.2).

3.4.3 Reconstruction and Tagging of Jets

In this Section we review software tools which are not part of the standard CDF production. These tools are relevant for the work presented in this thesis and will be often referred to in the next chapters.

Jets

The hadronisation of quarks produces several charged particles in form of jets. The number of particles produced and the transverse size of the jet is correlated to the energy of the quark. High energy jets detected in the calorimeters are widely used in High-p_T physics (top quark physics, Higgs searches). Jets produced in $b\bar{b}$ events typically have low energy and are reconstructed in the tracking system. Track-based jets with fixed cone size are formed with a simple algorithm, *Cone Clustering* [39]. The cone angle ΔR between two directions is generally defined as the distance in η-ϕ

space between the two directions

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.1}$$

An *isolation cone* ΔR_{iso} is defined with respect to a given direction, i.e. a reconstructed b-hadron momentum. The tracks falling into the isolation cone are not considered for the reconstruction of jets. At first the tracks with a minimum transverse momentum $p_{T,min}^{seed}$ are used as seed clusters and pairs of clusters with cone angle smaller than ΔR_{clm} are merged into a single cluster of momentum equal to the sum of the momenta of the seeds. The merging is repeated until no more merges are possible. Jets are formed by adding non-seed tracks with a minimum transverse momentum $p_{T,min}^{track}$ to an existing cluster if they form a cone angle smaller than ΔR_{jm} with the cluster axis. Every time a track is added to a cluster, the total momentum is recomputed to include the new track. ΔR_{iso}, ΔR_{clm}, ΔR_{jm}, $p_{T,min}^{seed}$ and $p_{T,min}^{track}$ are parameters of the algorithm.

b-jet Tagging

The tagging of b-jets is the main subject of this thesis. Some method devoted to b-jet tagging already exist and are based on the long lifetime of b-hadrons.

Secondary Vertex The *SecVtx* algorithm [40] has been developed for High-p_T physics. It searches within a jet for tracks which form a secondary vertex. The algorithm is run in two passes. First, well reconstructed tracks which satisfy a minimum cut on the impact parameter divided by its error, d_0/σ (*Impact Parameter Significance*), are selected. A secondary vertex fit is attempted with at least three of these tracks. If the fit is not successful, the cuts on the tracks are tightened and a vertex with at least two tracks is fitted in the second pass. The secondary vertex is accepted if its distance in the r-ϕ plane from the PV has large significance. The jet is then tagged as b-jet.
A secondary vertex fit succeeds only for a small fraction of jets in $b\bar{b}$ events. On the other hand, the jets tagged by the SecVtx algorithm are very often true b-jets. Further studies on the performance of this algorithm for $b\bar{b}$ events are presented in Section 5.6.1.

Jet Probability A second algorithm developed for High-p_T physics to tag b-jets, *JetProb* [41], does not explicitly search for a secondary vertex in the jet but evaluates the probability that the jet might contain one by combining the probability for each track to belong to the primary vertex. If a track belongs to a jet and it originating from a secondary vertex within the jet, the sign of the impact parameter with respect to the jet direction is expected to be positive (fig. 3.5). A probability, proportional to the likelihood that the track originates from the primary vertex, can be constructed from

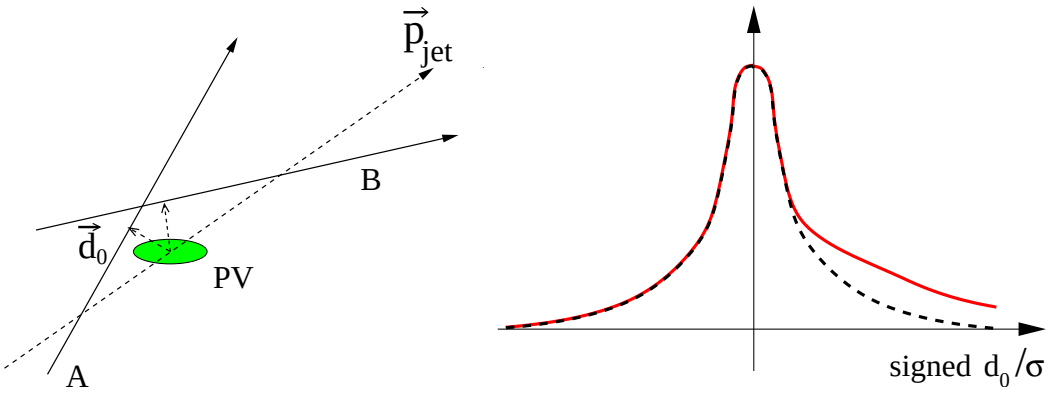

Figure 3.5: *Left:* Track A and B have similar impact parameters with respect to the primary vertex, $|\vec{d}_0|$, but opposite signed impact parameters, $d_0^{signed} = |\vec{d}_0| \cdot sign(\vec{d}_0 \cdot \vec{p}_{jet})$ *Right:* Distribution of signed impact parameter significance for tracks from the primary vertex (*dashed line*) and tracks from a secondary vertex (*solid line*).

the signed impact parameter significance. From a technical point of view it is equal to the probability for a track in the primary vertex to have an impact parameter at the measured value or larger. The probability distribution is flat for tracks originating from the primary vertex and it peaks at zero for tracks produced in the decay of long lived particles. The jet probability is given by

$$J_P = P(N) \cdot \sum_{n=1}^{N} \frac{(-logP(N))^n}{n!} \tag{3.2}$$

$$P(N) = \prod_{m=1}^{N} t_m \tag{3.3}$$

where N is the number of tracks in the jet and t_m is the probability of the m-th track. One of the developments presented in this thesis is a new method to compute the track and jet probability based on Neural Networks (Chapter 6). The track and jet probability provided by the JetProb algorithm serve as a benchmark for the new method.

Chapter 4

Outline of B_s^0 Oscillation Analysis

From the experimental point of view, the analysis of the B_s^0 oscillation corresponds to the study of the time dependent asymmetry between the number of events in which a generated B_s^0 decays as a B_s^0 (*unmixed events*, $N_{unmixed}$) and those in which it decays as a $\bar{B}_s^0$ (*mixed events*, N_{mixed})[1]. Using eqn. 2.9 and 2.10, the asymmetry is a function of the quantity we want to measure, Δm_s:

$$A(t) = \frac{N_{unmixed}(t) - N_{mixed}(t)}{N_{unmixed}(t) + N_{mixed}(t)} = \cos \Delta m_s \, t \tag{4.1}$$

$$N_{unmixed}(t) = N_0 \cdot \frac{e^{-t/\tau}}{2\tau} \left(1 + \cos \Delta m_s \, t\right) \tag{4.2}$$

$$N_{mixed}(t) = N_0 \cdot \frac{e^{-t/\tau}}{2\tau} \left(1 - \cos \Delta m_s \, t\right) \tag{4.3}$$

where $\tau = 1/\Gamma$ is the B_s^0 lifetime, t is the time in the B_s^0 rest frame and N_0 is the number of B_s^0 mesons at $t = 0$. A graph of $N_{unmixed}(t)$, $N_{mixed}(t)$ and $A(t)$ in the absence of detector effects is shown in fig. 4.1 for a B_d-like oscillation ($\Delta m_d = 0.5$ ps^{-1}) and for a B_s-like oscillation with a hypothetical $\Delta m_s = 15$ ps^{-1}.
Studies of the time-dependent asymmetry require three ingredients:

- the reconstruction of the signal in a flavour eigenstate
- the determination of the B_s^0 decay time
- the identification of the flavour of the B_s^0 at production and at decay time

[1]Here and in the following the charge conjugate modes are implicitly included, unless explicitly stated otherwise.

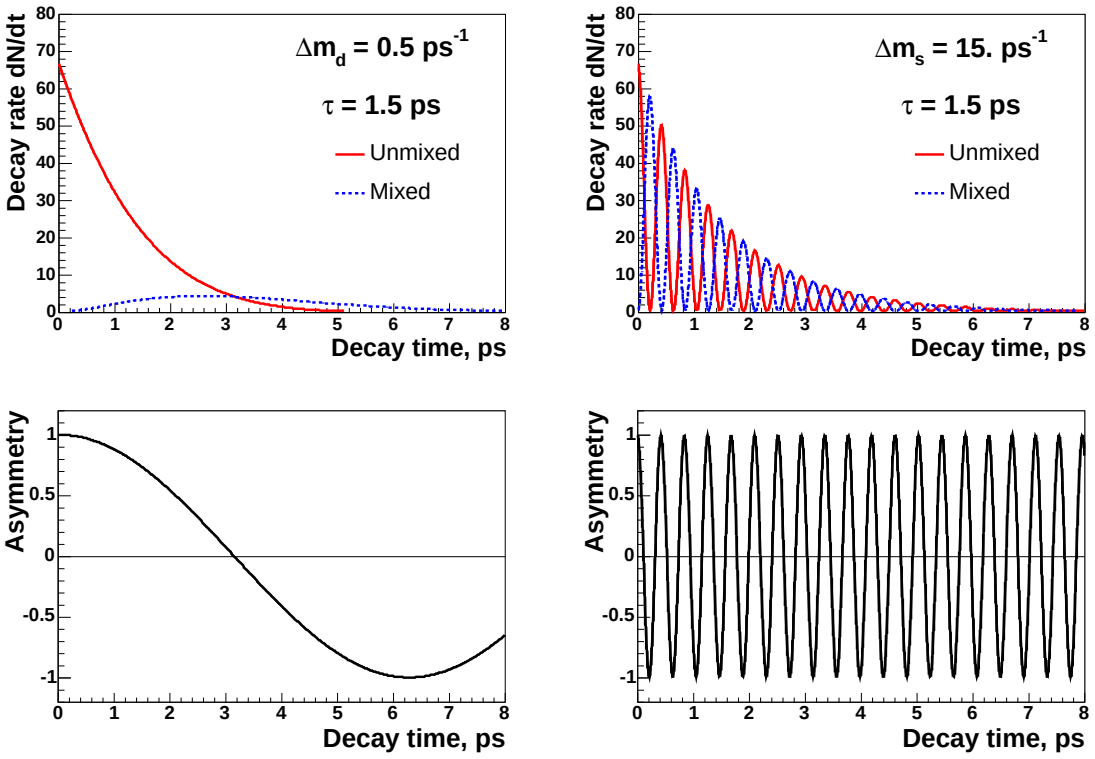

Figure 4.1: *Top:* Number of mixed and unmixed events as a function of the decay time for a B_d-like oscillation (*left*) and a faster B_s-like oscillation (*right*). *Bottom:* Asymmetry for a B_d-like oscillation (*left*) and a B_s-like oscillation (*right*). The frequency $\Delta m_s = 15$ ps^{-1} is hypothetical. Detector effects are not taken into account.

The significance for a B_s^0 oscillation signal can be approximated by the expression

$$Sig = \sqrt{\frac{S\epsilon D^2}{2}} \sqrt{\frac{S}{S+B}}\, e^{-\frac{(\Delta m_s \sigma_t)^2}{2}} \tag{4.4}$$

where S is the number of reconstructed B_s^0 candidates, ϵD^2 is the tagging power (see Section 4.3.1 for details), B is the number of background events and σ_t is the resolution on the proper time measurement. The proper time resolution depends on the decay length resolution and the momentum resolution. The significance decreases rapidly when Δm_s increases due to the exponential factor, which also depends on the proper time resolution. The ability to resolve rapid B_s^0 oscillations thus requires excellent decay length and momentum resolution, and benefits from having a low mistag rate and a high B_s^0 purity.

This chapter presents the building blocks needed for an oscillation analysis, referring

in particular to the analysis implemented by CDF for the Winter Conferences 2005. Figure 4.2 shows a sketch of a B_s^0 event at the Tevatron. The b-quarks are produced in pairs[2] and fragment independently into b-hadrons. The B_s^0 is reconstructed by combining its decay products in a vertex (see Section 4.1). As already mentioned, B decay vertices are displaced with respect to the primary vertex. Consequently displaced tracks are a signature for a b-hadron decay. The proper decay time (Section 4.2) is computed from the transverse momentum and the decay length in the transverse plane, L_{xy}. The latter is defined as the projection of the distance between the primary vertex and the decay vertex onto the B_s^0 transverse momentum. The tracks in the vicinity of the B_s^0 define the *same side* of the event. The other tracks belong to the *opposite side* and are employed to identify the second b-hadron produced in the event. The B_s^0 flavour at decay time can be determined from the decay products. The flavour at production time can be determined with particles accompanying the fragmentation $\bar{b} \to B_s^0$ (*fragmentation tracks*), or with the decay products and fragmentation tracks from the opposite side b-hadron (Section 4.3). The work presented in this thesis focuses on this particular step of the mixing analysis. The combination of the ingredients in a maximum likelihood fit based on the mixing probability (eqn. 4.2 and 4.3) to extract Δm_s is described in Section 4.4, where the latest results from CDF are presented as well.

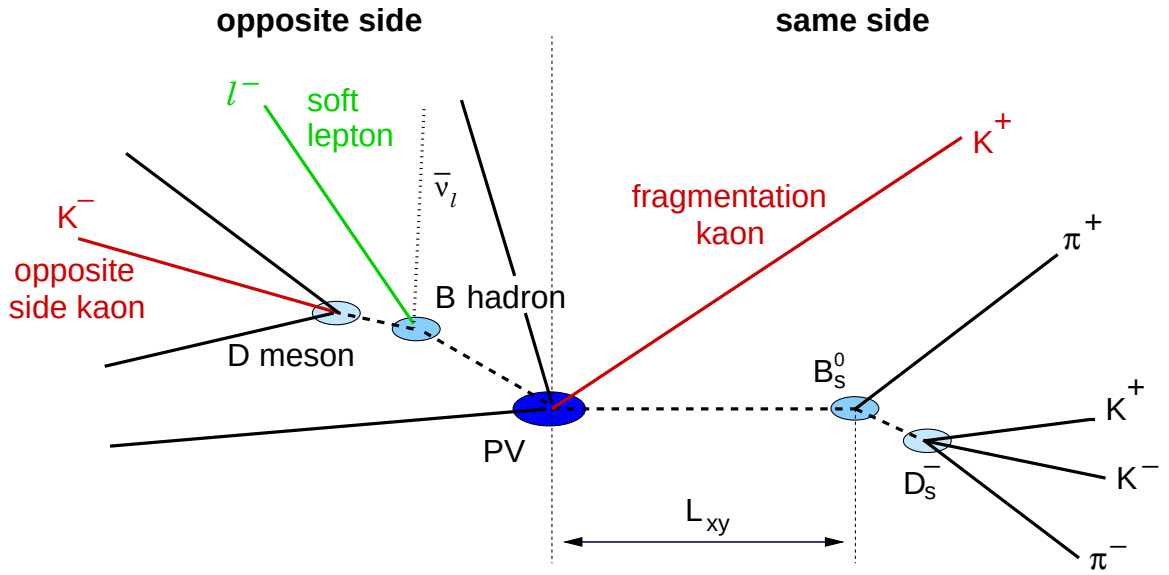

Figure 4.2: Schematic view of a B_s^0 event at the Tevatron summarising the elements needed for the B_s^0 mixing analysis.

[2]More details about $b\bar{b}$ production at the Tevatron are given in Section 5.1.

4.1 Signal Reconstruction

The B_s^0 signal can be reconstructed by combining some or all the products of its decay in a vertex. Various oscillation analyses differ mostly in the way the signal is reconstructed, which in turn affects the quality of the decay flavour tag and the B_s^0 purity. Some analyses perform an inclusive reconstruction, that identifies only part of the B_s^0 decay products, and others reconstruct the signal exclusively, finding all the particles generated in the B_s^0 decay. Inclusive methods have the advantage of large statistics, but suffer from low purity and worse momentum resolution. Exclusive methods yield small sample sizes but benefit from a much increased resolution on momentum and decay length per event (eqn. 4.4). In the CDF mixing analyses [42, 43] the B_s^0 have been reconstructed in modes including a D_s^- because the branching fraction for

$$B_s^0 \to D_s^- X \tag{4.5}$$

is $(94 \pm 30)\%$ [1]. Therefore D_s^- mesons appear frequently in B_s^0 events. The decay channels considered for the D_s^- are:

- $D_s^- \to \phi\, \pi^-$ (3.6 ± 0.9 %), $\phi \to K^+K^-$ (49.1 ± 0.6 %)
- $D_s^- \to K^{*0}\, K^-$ (3.3 ± 0.9 %), $K^{*0} \to K^\pm\pi^\mp$ ($\sim 100\%$)
- $D_s^- \to \pi^-\, \pi^\pm\, \pi^\mp$ (1.01 ± 0.28 %)

where the numbers in brackets indicate the branching ratio for the decay mode [1]. The background component in each reconstructed channel has been estimated studying the mass and lifetime distributions of possible sources of mis-reconstructed B_s^0 mesons in simulated events. A fit to the B_s^0 mass and lifetime distribution has been performed with a floating background fraction to determine the signal purity in data.

4.1.1 Semileptonic Channel

When the B_s^0 undergoes a *semileptonic decay*, i.e. it decays in

$$B_s^0 \to \ell^+ \nu_\ell X$$

the neutrino escapes detection. Thus the B_s^0 can only be reconstructed inclusively. Due to the missing ν energy, the B_s^0 mass peak is very broad.
CDF has performed a mixing analysis [42] on a sample of semileptonic decays

$$B_s^0 \to D_s^-\, \ell^+\, \nu_\ell \tag{4.6}$$

reconstructed in the ℓ+SVT dataset (see Section 3.3.2). In order to identify the signal and understand the sources of background involving real D_s mesons the charge correlation between the lepton and the D_s has been used. The wrong sign combination

$(\ell^- D_s^-)$ corresponds to background events. Combinatorial background has been studied by means of the D_s^- sidebands. The selection for each channel has been optimised in order to maximise $S/\sqrt{S+B}$, which is one of the factors that contributes to the B_s^0 oscillation significance (eqn. 4.4). The D_s mass peak in the $\phi\pi$ channel is shown in fig. 4.3 (left). This channel is particularly pure due to the clean ϕ signature. About 7500 B_s^0 candidates were found in total.

4.1.2 Hadronic Channel

When all the decay products enter the vertex reconstruction, the B_s^0 is reconstructed exclusively. This is only possible when all the final state particles are charged hadrons. A sharp B_s^0 mass peak is then observed and the events in the peak form a high purity B_s^0 signal.
A mixing analysis [43] has been performed at CDF by reconstructing in the TTT sample (see Section 3.3.1) B_s^0 mesons exclusively in the channel[3]

$$B_s^0 \to D_s^- \, \pi_B^+ \tag{4.7}$$

The events are triggered by two displaced tracks. As in the semileptonic channel, the selection of the B_s^0 candidates is tuned to maximise $S/\sqrt{S+B}$. The mass peak for the candidates in the $\phi\,\pi\,\pi$ channel is shown in fig. 4.3 (right). The total number of B_s^0 candidates selected in the TTT sample is about 900.

4.2 Proper Time Reconstruction

The asymmetry in eqn. 4.1 depends on the time in the B_s^0 rest frame t, also called *proper time*, which is calculated as follows for the instant at which the B_s^0 decays:

$$t = \frac{L_{xy}(B_s) \cdot m(B_s)}{c \cdot p_T(B_s)} \tag{4.8}$$

where $L_{xy}(B_s)$ is the decay length, $m(B_s)$ is the mass of the b-hadron, $p_T(B_s)$ is the transverse momentum of the b-hadron and c is the speed of light.
For an inclusively reconstructed B_s^0, a correction factor is needed in eqn. 4.8 to take into account that the p_T and L_{xy} of the B_s^0 are estimated using only some of the decay products. Such a correction, called *k-factor*, is evaluated by comparing the observed momentum with the true p_T of the B_s^0 in simulated events. For semileptonic decays, e.g. $B_s^0 \to D_s^- \ell^+ \nu_\ell$, the k-factor is

$$k = \left\langle \frac{p_T(\ell D_s)}{p_T(B_s)_{\rm MC}} \frac{L_{xy}(B_s)_{\rm MC}}{L_{xy}(\ell D_s)} \right\rangle \tag{4.9}$$

[3]The notation π_B is meant to distinguish the pion from the B_s^0 decay from those of the D_s decay.

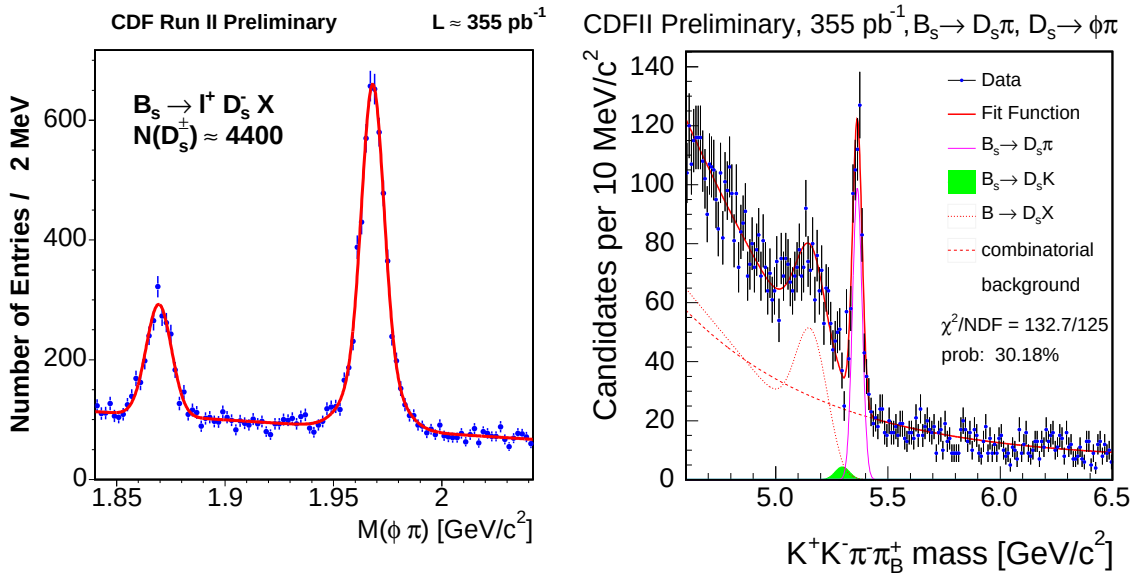

Figure 4.3: *Left:* D_s^- mass peak in the $\phi\pi$ decay channel in the semileptonic sample [42]. The peak on the left corresponds to D^- mesons in Cabibbo suppressed channels; the peak on the right is centred at the D_s^- mass. *Right:* a B_s^0 mass peak in the $\phi\,\pi\,\pi$ channel in the hadronic sample [43]. The candidates in the low sideband are $B_s^0 \to D_s^{*-}\pi_B^+$, $D_s^{*-} \to D_s^- X$.

This definition yields a distribution rather than a single number, which depends on the measured quantities. The proper time is then convoluted with the k-factor distribution:

$$t = \frac{L_{xy}(\ell D_S) \cdot m(B_s)}{c \cdot p_T(\ell D_s)} \otimes k \tag{4.10}$$

The introduction of the k-factor causes a smearing of the proper time distribution which affects the resolution σ_t.

Trigger requirements or selection cuts may affect the proper decay time distribution as well. This bias needs to be taken into account and can in general be described by an appropriate ct-efficiency function. The SVT tracks have to pass a cut on the impact parameter (Section 3.3). When these tracks are used to reconstruct the B_s^0 this translates into a bias of the decay length of the vertex. To give a simple representation of the effect of a lower and upper cut on L_{xy}, two step functions have been convoluted with the distributions of $N_{unmixed}(t)$ and $N_{mixed}(t)$ in fig. 4.4 for a B_s-like oscillation. Most of the B_s^0 candidates with early decay times are lost. The asymmetry graph is not affected for small values of the decay time, although the first and the last oscillation periods have to be evaluated with poorer statistics than in the ideal case (fig. 4.1) and consequently have much larger errors. The effect on the B_d oscillation graph is negligible and thus it is not shown.

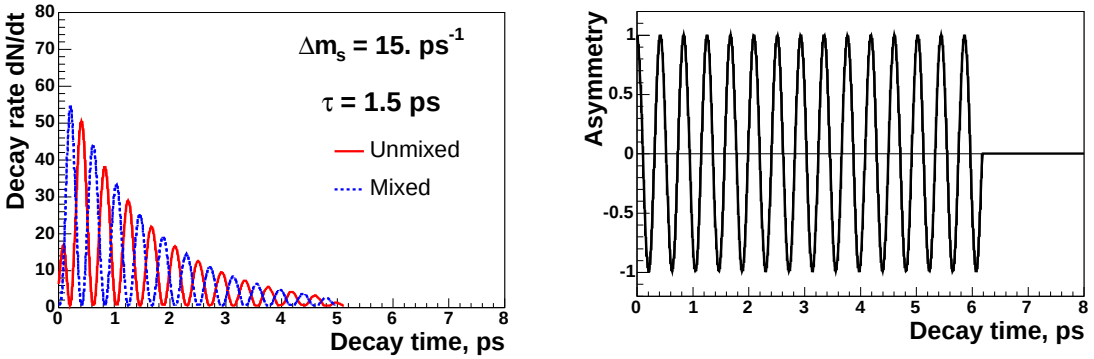

Figure 4.4: Effect on the asymmetry graphs of a lower and an upper cut on the decay length for a B_s-like oscillation. The decay rate graph has been convoluted with a step function at $t = 0.1\,\mathrm{ps}$ and at $t = 5\,\mathrm{ps}$, which corresponds to a cut $30\,\mu\mathrm{m} < \mathrm{L_{xy}} < 1.5\,\mathrm{mm}$ for a B_s^0 with transverse momentum of 5 GeV/c.

The proper time resolution is given by

$$\sigma_t = \sqrt{\sigma_{vtx}^2 + \left(t \cdot \frac{\sigma_p}{p}\right)^2} \tag{4.11}$$

The first contribution, σ_{vtx}, gives a time-independent contribution and depends on the resolution of L_{xy}, i.e. on the reconstruction of the primary and secondary vertex. The second one, σ_p/p, is mainly determined by the k-factor for inclusive channels and its contribution grows with t. The effect of the proper time resolution is a damping of the oscillation by a factor $exp\,(-(\Delta m_s \sigma_t)^2/2)$ and is illustrated in fig. 4.5 for the B_s^0 candidates reconstructed exclusively and inclusively. The asymmetry graph for B_d is barely affected and it is not shown. The effect on the B_s-like asymmetry graph is dramatic, especially for the inclusively reconstructed candidates. Clearly, the events at low t dominate the asymmetry measurement, but due to the trigger bias and to the selection cuts to ensure a low background fraction, they are unfortunately the most difficult to identify.

The proper time distribution for the exclusive B_s^0 candidates in the $\phi\,\pi\,\pi$ channel reconstructed on data is shown in fig. 4.6.

4.3 b-Flavour Tagging

b-Flavour Tagging is a collective term for several experimental techniques to find out whether a hadron contains a b- or a $\bar{b}$-quark. The techniques are mostly based on

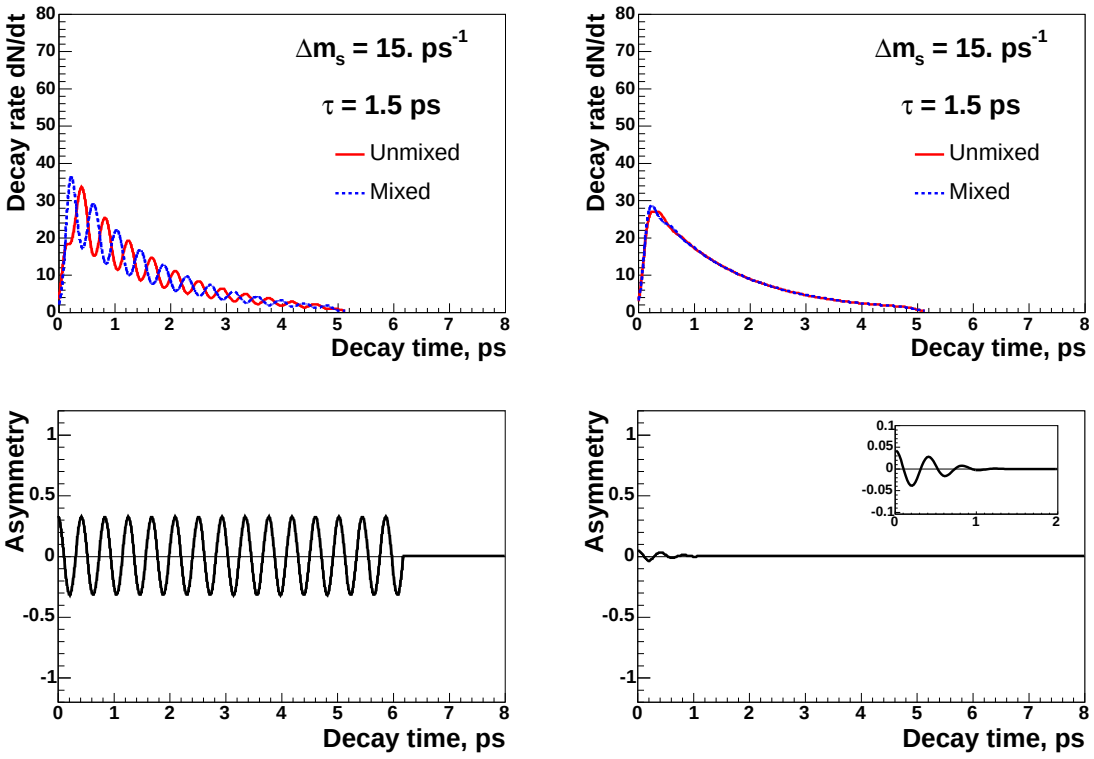

Figure 4.5: Effect of the proper time resolution σ_t on the asymmetry graphs for an exclusive channel (*left*) and an inclusive channel (*right*) for a B_s-like oscillation. For the graph relative to the exclusive channel the values $\sigma_{vtx} = 30\mu m$ (100 fs) and $\sigma_p/p = 0$ have been used. The graph representing the effect on the inclusively reconstructed channel has been produced with the values $\sigma_{vtx} = 50\mu m$ (167 fs) and $\sigma_p/p = 15\%\,(\mathrm{GeV/c})^{-1}$

charge correlation between the b-quark and tracks in the event. Thus the tagging of the b-quark corresponds to the determination of its charge. There are different kinds of taggers

Same Side Taggers use the products of the fragmentation of the reconstructed B

Opposite Side Taggers exploit the information from the other b-quark in the event to determine the flavour of the reconstructed B

The tagging methods are reviewed in Sections 4.3.2 and 4.3.3.

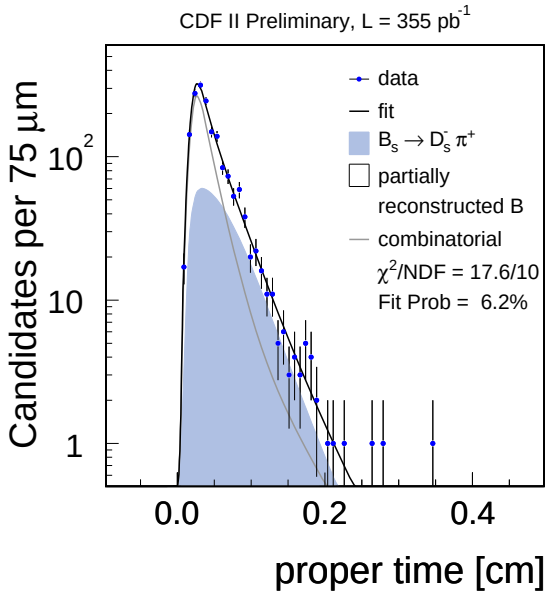

Figure 4.6: Proper time measurement for exclusively reconstructed B_s^0 mesons in the $\phi\,\pi\,\pi$ channel [43].

4.3.1 Tagging Power ϵD^2

Assuming that the flavour of a b-hadron is known, a tagger gives a *Right Sign Tag* (RS) if its decision agrees with the b-hadron flavour, and a *Wrong Sign Tag* (or *mistag*, WS) if the decision disagrees. If the tagger is not able to give a decision, the b-hadron is *Non Tagged* (NT).
The tagger *efficiency* is given by

$$\epsilon = \frac{N_{RS} + N_{WS}}{N_{RS} + N_{WS} + N_{NT}} \tag{4.12}$$

where N_{RS} is the number of the right sign tags, N_{WS} is the number of the wrong sign tags and N_{NT} is the number of not tagged events in the sample.
The *dilution* of a tagger is defined as the asymmetry between right and wrong tags:

$$D = \frac{N_{RS} - N_{WS}}{N_{RS} + N_{WS}} \tag{4.13}$$

The dilution is related to the probability for the tagger to give a right sign tag, P_{RS}, or a mistag, P_{WS}:

$$P_{RS} = \frac{N_{RS}}{N_{RS} + N_{WS}} = \frac{1 + D}{2} \tag{4.14}$$

$$P_{WS} = \frac{N_{WS}}{N_{RS} + N_{WS}} = \frac{1 - D}{2} \tag{4.15}$$

The name "dilution" is therefore misleading, as good tagging algorithms have high dilution and poor taggers have dilution close to zero. The use of dilution instead of the probability measure for a correct tag is justified when the statistical power of the flavour tag is examined. The relation between the true asymmetry in the data and the measured asymmetry is

$$A_{true} = \frac{1}{D} A_{meas} \tag{4.16}$$

which is derived in Appendix A. The statistical uncertainty on A_{true} is expressed by

$$\sigma_A = \sqrt{\frac{1 - D^2 A^2}{\epsilon D^2 N}} \tag{4.17}$$

where N is the number of events before tagging. The uncertainty scales like $1/\sqrt{\epsilon D^2}$, thus the statistical power of a tagger is quantified by ϵD^2. Consequently, optimising ϵD^2 is the primary goal when developing a flavour tagger. The asymmetry amplitude (fig. 4.5), as eqn. 4.16 indicates, is further reduced by a factor D, which has typical values of few percent, thus the mistag rate adds another complication to the oscillation analysis.

It proves helpful for mixing analyses to study the dependency of the tagger dilution on event-shape variables. A parameterisation of D as a function of such variables allows to compute the expected dilution in each event, which is then included as a weight in the oscillation fit

$$N_{unmixed}(t) \propto e^{-t/\tau} \left(1 + D \cos \Delta m_s \, t\right) \tag{4.18}$$

$$N_{mixed}(t) \quad \propto e^{-t/\tau} \left(1 - D \cos \Delta m_s \, t\right) \tag{4.19}$$

As a result, events with clear indications of mixing are weighted more than events for which no strong decision can be taken and this improves the significance of the oscillation measurement.

The power of a tagger has to be measured on a calibration sample, i.e. a sample of b-hadrons for which the flavour is known (e.g. charged b-hadrons or a sample not enriched in neutral B-mesons). The extraction of the efficiency and the dilution, and eventually the determination of a parameterisation of the dilution from the data, is called *Tagger Calibration*. Each CDF tagger has an own calibration procedure. The calibration of the Jet Charge Tagger presented in this thesis is described in detail in Section 7.3.

4.3.2 Same Side Tagging

Kaon Tagging In the specific case of B_s mesons, kaons can usually be found among the fragmentation products. When a $\bar{B}_s^0$ meson is produced, a $s\bar{s}$ pair is extracted

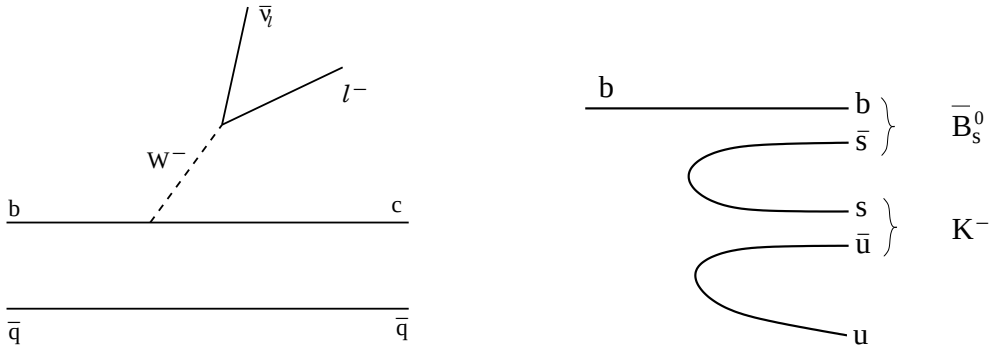

Figure 4.7: *Left:* Diagram of the semileptonic decay of a B meson. The charge correlation between the lepton and the heavy quark is the basis of the Soft Lepton Tagger. *Right:* Production of a charged K^- in association with a $\bar{B}_s^0$ meson in the fragmentation of a b-quark.

from the vacuum (fig. 4.7, right). The $\bar{s}$ quark couples with the b-quark to form the $\bar{B}_s^0$. The s quark can associate with other quarks and if it associates with a $\bar{u}$ then a K^- is produced. Positively charged kaons can be produced together with B_s^0 mesons. This correlation is exploited in the Same Side Kaon Tagger, which gives information on the flavour of the B_s at production time. Similarly, B_d^0 mesons can be tagged using fragmentation pions. In addition to the difficult selection of the right particle among all of the fragmentation tracks, the correct identification of kaons is not trivial, therefore this tagger has to rely on very good particle identification.
At CDF Same Side Kaon Tagging has been studied and it can be applied to B_s in a mixing analysis. The measurement of the tagging power on data will only be possible once the B_s oscillation frequency has been measured [44]. The development of the Same Side Kaon Tagger and its use for setting a limit on the mixing frequency requires good understanding of the fragmentation processes and reliable simulation.

4.3.3 Opposite Side Tagging

Kaon Tagging Due to the dominant quark-level decay$b \to c \to s$ kaons can often be found among the decay products of b-hadrons. Negatively (positively) charged kaons are produced in the decays of b-hadrons containing the b- ($\bar{b}$-) quark. The charge of the kaon provides the tag of the b-hadron flavour. The tag is wrong when the opposite side b-hadron is a mixed neutral b-meson. About 17% of the neutral b-mesons mix before decaying [1].
At CDF Opposite Side Kaon Tagging is not feasible at the moment due to the difficult inclusive selection of kaons from the b-hadron decay chain.

Soft Lepton Tagging The Soft Lepton Tagger exploits the charge correlation between the b-quark and the lepton produced in the semileptonic decay (fig. 4.7, left), where the lepton is either an electron or a muon. The b- ($\bar{b}$-) quark is associated with a negatively (positively) charged lepton, thus the charge of the lepton gives directly the flavour of the b-quark. About 20% of the b-hadrons decay semileptonically with either an electron or a muon in the final state. The measurement of the b-flavour is wrong when the lepton stems from the sub-sequential D-meson decay

$$b \to c \ \to s\, \ell^+\, \nu_\ell \tag{4.20}$$

Similarly to the Kaon Tagging case, mixed neutral b-mesons give a wrong tag as well. The Soft Lepton Tagger achieves high purity but the efficiency is rather low due to the relatively low fraction of semileptonic decays of the b-hadrons and the difficult identification of leptons at low momentum.
At CDF the soft lepton tagging is performed by combining variables sensitive to lepton identification in a likelihood ratio [45].

Jet Charge Tagger An empirical correlation between the charge of the tracks in the vicinity of the b-quark and the b-quark flavour has been observed [46]. The flavour information is carried by both, fragmentation tracks and the b-hadron decay products. It can be extracted with a weighted sum of the tracks' charge. Jets are reconstructed on the opposite side to identify the opposite side b-hadron inclusively and the tracks in the jet that is most likely to contain the b-hadron decay products are used to compute the weighted jet charge.
For the CDF mixing analyses presented at the Winter Conferences 2005 [2] a cut based Jet Charge Tagger has been used [47, 48]. This tagging algorithm searches for a jet on the opposite side with a secondary vertex (see Section 3.4.3); if no jet with a secondary vertex is found, a jet with low probability to come from the primary vertex (see Section 3.4.3) is searched for; if neither a low probability jet is found, the jet with the highest transverse momentum on the opposite side is selected.
In general, the Jet Charge Tagger is characterised by low purity and high efficiency. The tagging power of the Soft Lepton Taggers and of the Jet Charge Tagger is listed in table 4.1. The Jet Charge Tagger is as effective as the Soft Muon Tagger, but it has a ten times larger efficiency. Consequently, a large tagging power gain can come from its improvement.
The *acceptance*, i.e. the fraction of events in which both the b-quarks are in the η-coverage of the detector, affects the dilution of this tagger. Moreover, $b\bar{b}$ events in the p_T range observed at the Tevatron often do not display a clear jet structure. It follows that the identification of the b-jet is the major issue for this tagger. The work described in this thesis is mainly devoted to the understanding and the solution of the problem of b-jet selection. Relevant features of $b\bar{b}$ events are described in Chapter 5.

Opposite Side Tagger	ϵ, %	ϵD^2, %
Soft Electron	~ 3	0.366 ± 0.031
Soft Muon	~ 5	0.698 ± 0.042
Jet Charge	~ 82	0.715 ± 0.027

Table 4.1: Tagger effectiveness for the opposite side taggers applied in the CDF mixing analyses. The ϵD^2 has been calibrated on the ℓ+SVT sample.

The strategies considered for the identification of b-jets and the resulting Jet Charge Tagger algorithm for CDF are described in Chapter 6 and in Chapter 7, respectively.

The flavour tagging in CDF mixing analyses has been performed with the Jet Charge and the Soft Lepton Tagger. The Same Side Kaon Tagger has not been included for the first iteration of the analysis, but is going to be considered for the next measurement. The combination of several correlated taggers is supposed to be more powerful than a single tagger. Nevertheless, the CDF mixing analysis applied the Jet Charge Tagger and the Soft Lepton Tagger in an exclusive fashion, giving priority to the tagger with the higher dilution with the aim of eliminating the correlations between the taggers. Further studies of the correlation of taggers are needed to allow a meaningful combination. For the next update of the mixing analysis it is planned to use a combination of Same Side Kaon Tagging and Opposite Side Taggers.

4.3.4 Remarks with Respect to B-Factories

The application of Soft Lepton and Jet Charge Tagger to the CDF B_s^0 mixing analysis scores about 1.4% for ϵD^2. This number is rather small, especially in comparison to typical tagging powers at the B-Factories, which are of the order of 20 - 30% for Soft Lepton Tagging. It is important to stress that b-flavour tagging at the Tevatron is extremely challenging. The B-Factories produce the resonance $\Upsilon(4S)$, which decays into B mesons. The decay products are quantum correlated, thus they evolve coherently and conserve their flavour anti-correlation. As soon as one of them decays in the instant t_0, the other has the probability to mix given by eqn. 2.9 and 2.10 with the substitution of $t - t_0$ to t. The flavour at decay time t_0 of one of them tags then automatically the flavour of the other at the instant t_0. In addition, the neutral mesons are produced pairwise in decays of the $\Upsilon(4S)$, so no other particles are present in the detector besides the final state products of the B^0 and $\bar{B}^0$. The situation is completely different at the Tevatron, where the hadronic environment adds some complications: beam remnants and fragmentation tracks populate the events together with tracks from interesting decays, summing up to an average number of detected charged particles of the order of 50 for $b\bar{b}$ events. Consequently, the extraction of the signal is not

easy and the identification of a particular fragmentation particle is extremely difficult. It is necessary to keep in mind the complexity of $b\bar{b}$ events at the Tevatron, especially when dealing with b-flavour tagging.

4.4 Extraction of Δm_s

The mixing frequency of B_s^0-$\bar{B}_s^0$ mesons has not yet been measured, thus, unlike Δm_d, it is not possible to extract Δm_s directly from a fit of the asymmetry $A(t)$ because there is no sensible starting value for the frequency parameter in the fit. Instead the range of possible Δm_s values is probed with the *amplitude scan* method [49]. An extra parameter, the *amplitude* $\mathcal{A}$, is included to the expression of the oscillation probability (eqn. 4.2 and 4.3)

$$N_{unmixed}(t) \propto e^{-t/\tau}\,(1 + \mathcal{A} \cdot D \cos \Delta m_s \cdot t) \tag{4.21}$$

$$N_{mixed}(t) \propto e^{-t/\tau}\,(1 - \mathcal{A} \cdot D \cos \Delta m_s \cdot t) \tag{4.22}$$

The asymmetry in the data is fit for fixed frequency values. The amplitude is left free to float during the fit. In the case of infinite statistics, optimal resolution and perfect tagger parameterisation, $\mathcal{A}$ is expected to be unit for the true oscillation frequency and zero for the remaining of the probed spectrum. In real life, the output of the procedure is a list of fitted values $(\mathcal{A}, \sigma_{\mathcal{A}})$ for each Δm_s hypothesis. Such a Δm_s hypothesis is excluded to a 95% confidence level in case the following relation is observed

$$\mathcal{A}(\Delta m_s) + 1.645 \cdot \sigma_{\mathcal{A}}(\Delta m_s) < 1 \tag{4.23}$$

The sensitivity at 95% confidence level is determined by the value of σ_A which satisfies the relation

$$1.645 \cdot \sigma_{\mathcal{A}}(\Delta m_s) = 1 \tag{4.24}$$

In order to use the amplitude scan method, all the parameters in the fit except for $\mathcal{A}$ have to be fixed.

A sample of B_d^0 and B^+ candidates has been used to fine-tune the calibration of the dilution of each tagger [50] (see Section 4.3.1). The kinematics of the events selected for the B_s^0 mixing analysis might be different from the kinematics of the events in the calibration sample. This might affect the dilution measurement, therefore a re-calibration is needed. The B_d^0 sample is also a good test-bed for the overall fitting procedure. Figure 4.8 shows the result of the amplitude scan on the B_d^0 sample. The amplitude is compatible with zero over all the Δm_d spectrum, except for an interval in the vicinity of the true Δm_d value.

The result of the CDF amplitude scan on the semileptonic B_s sample is shown in fig. 4.9 (left). An analogous scan has been performed on the hadronic sample and the

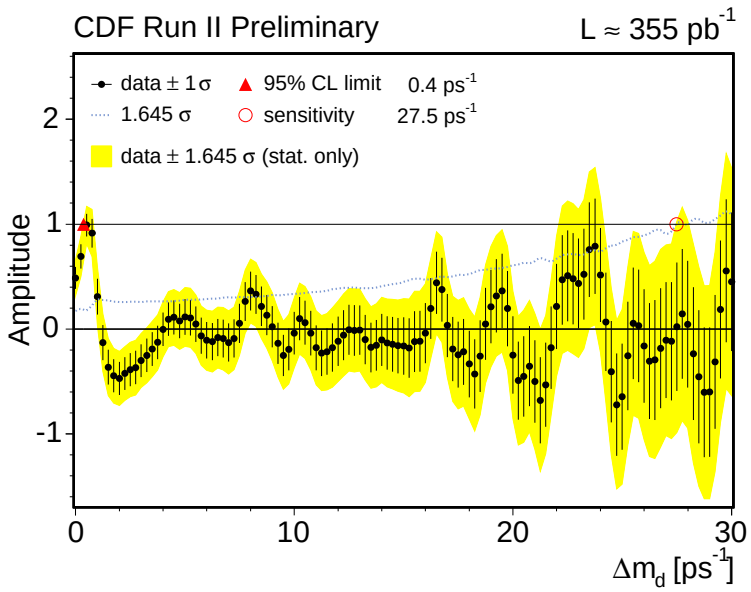

Figure 4.8: Amplitude scan for B_d mixing at CDF. The measurement of Δm_d is a test for the fitting procedure and offers the calibration for the dilution of the taggers [50].

result is illustrated in fig. 4.9 (right). The combined sensitivity of the semileptonic and the hadronic analysis is 8.4 ps^{-1} and the limit is

$$\Delta m_s > 7.9 \text{ ps}^{-1} \tag{4.25}$$

with 95% confidence level [2]. This constitutes the first limit by CDF on Δm_s with Run II data and it has been published for the Winter Conferences of 2005.
The combination of the CDF result with the Δm_s results of SLD and LEP experiments is shown in fig. 4.10 and corresponds to the exclusion of $\Delta m_s < 14.5$ ps^{-1} and a sensitivity of 18.6 ps^{-1} with 95% confidence level. The graph in fig. 4.11 shows the amplitude for $\Delta m_s = 15\,\text{ps}^{-1}$ measured so far by different experiments, including the latest CDF and DØ results. The precision achieved by CDF is better than the precision of the DØ measurement. The combined amplitude is compatible with zero within two sigma confidence, therefore this value of Δm_s can be excluded.
The next update of the CDF oscillation analysis, which is going to be published at the Summer Conferences 2005, will benefit from several improvements. Due to the increased luminosity delivered by the Tevatron and to the higher number of reconstructed decay modes, the size of the B_s^0 sample will be larger. At the same time work is going on to improve the vertex reconstruction, and thus the resolution on the proper time. The inclusion of the Same Side Kaon Tagger is foreseen in combination with Opposite Side Taggers. The Neural Network based Jet Charge Tagger, which is part of the work presented in this thesis, is going to substitute the cut based Jet Charge Tagger and an improvement in the tagging power is demonstrated in this thesis.

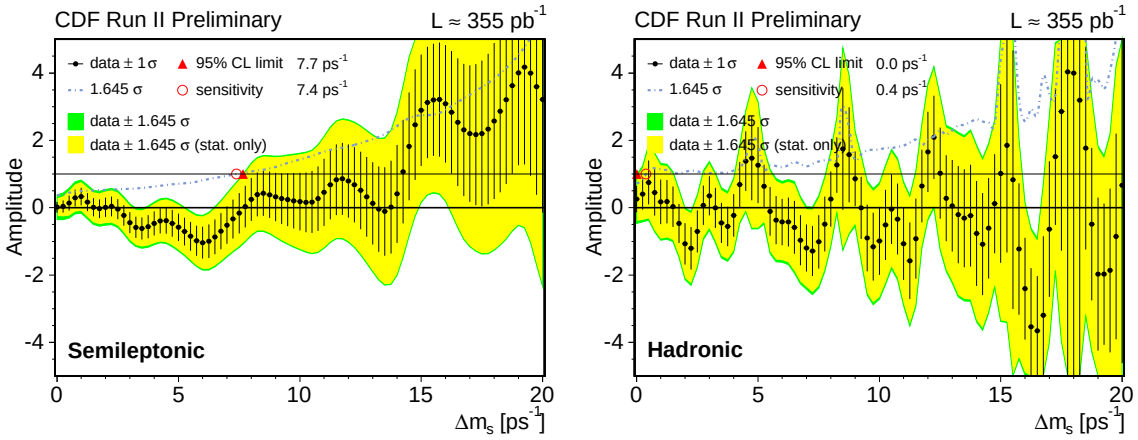

Figure 4.9: Amplitude scan for B_s mixing frequency at CDF in the semileptonic channels [42] (*left*) and in the hadronic channel [43] (*right*).

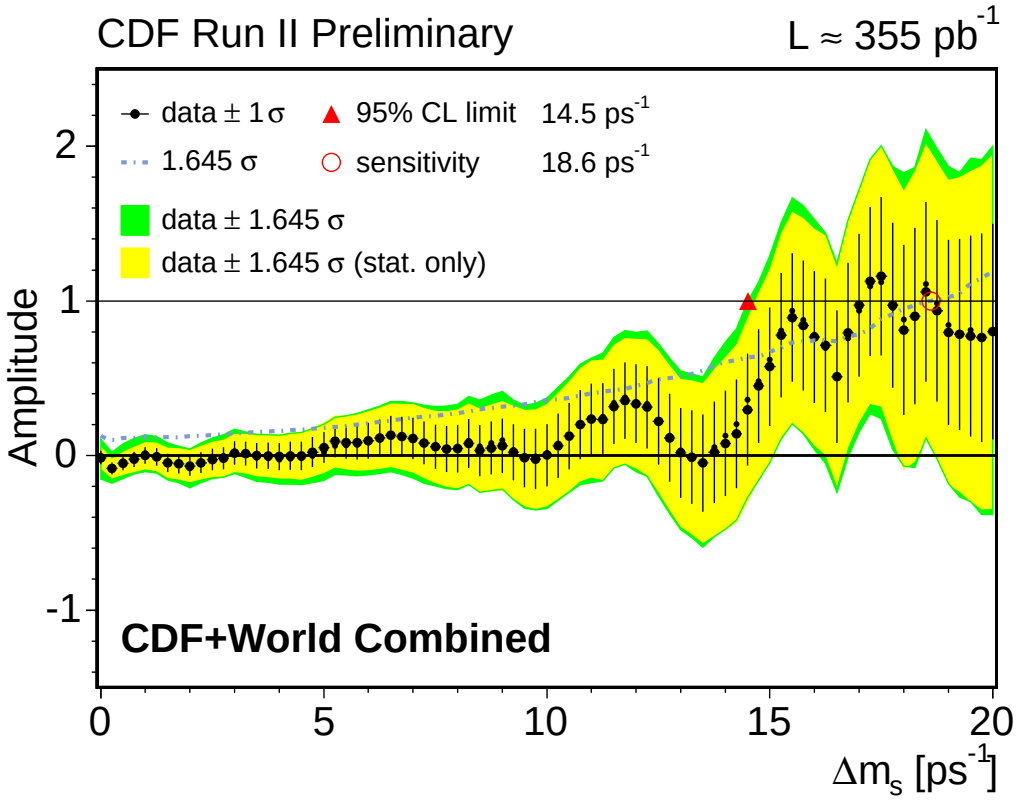

Figure 4.10: CDF result for Δm_s combined with the previously existing measurements [2].

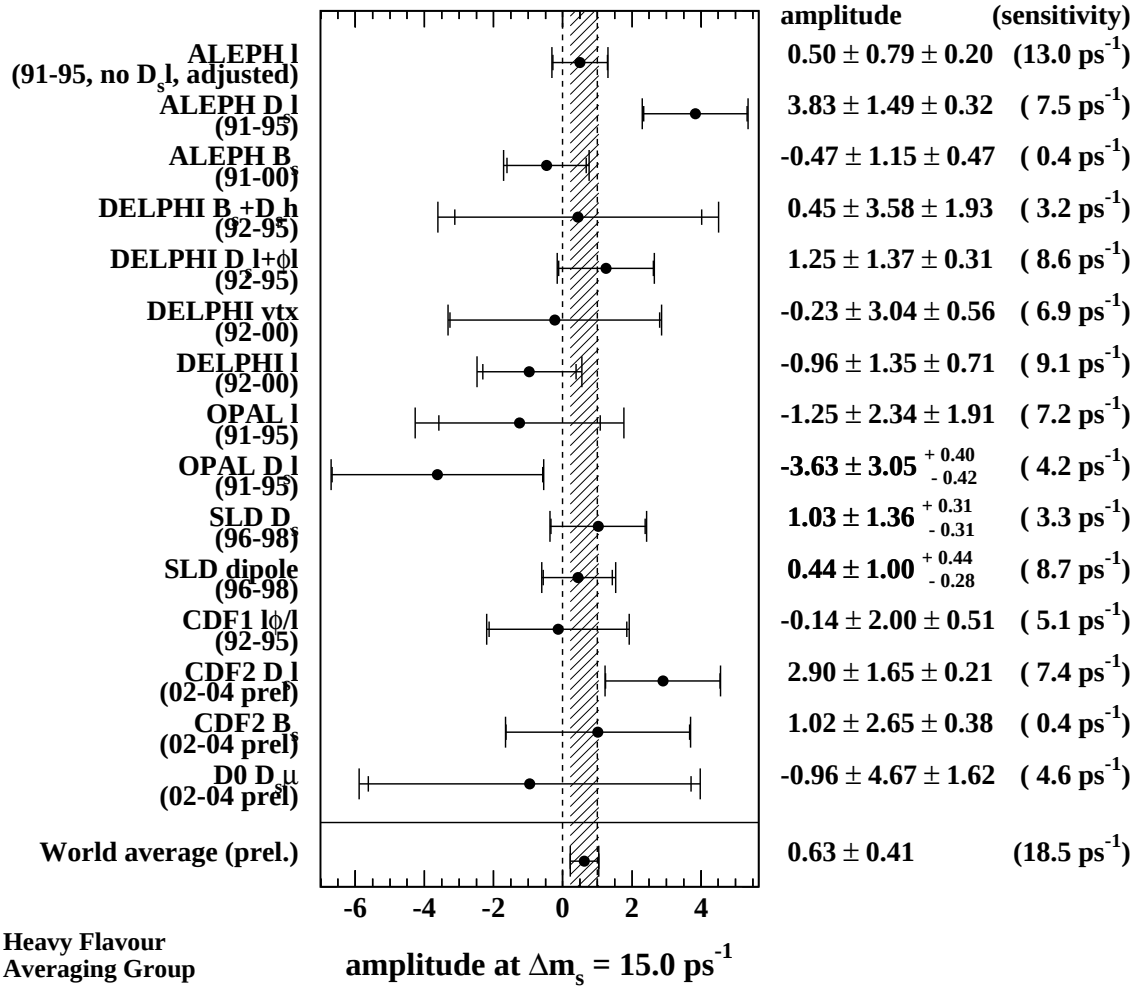

Figure 4.11: Summary of the amplitude measurements for $\Delta m_s = 15\,\mathrm{ps}^{-1}$ [51].

Chapter 5

Understanding $b\bar{b}$ Events with Simulation

In this chapter the features of events triggered by a lepton and a displaced track at CDF are studied on data and on Monte Carlo. The definition of event sides, their characteristics and their correlations are investigated.
The Monte Carlo generated with Pythia [54] used up to now in the CDF B group contains only leading-order $b\bar{b}$ pair production. It has turned out that this sample does not describe the features of the opposite side accurately. Additional $b\bar{b}$ pair creation processes had to be taken into account to describe the data.
In this chapter an alternative sample composition, including more $b\bar{b}$ creation processes, is proposed and it is compared to the previous sample. It is demonstrated that the alternative sample gives a better description of the data. It is therefore used in the following to study the characteristics of the event which are relevant for a flavour tagger, e.g. the correlation between the heavy quarks, the angular distribution of tracks from the decay of the b-hadrons and fragmentation tracks, the relation between the event sides. The results emphasise how difficult the task of tagging b-jet can be. However they can help to find the best way to identify the decay products of the opposite side b-hadron with jet clustering algorithms.
The study presented in this thesis is the first one within the CDF II B group which has taken the advantage of a Monte Carlo sample with additional processes.

5.1 $b\bar{b}$ Pair Production at Tevatron

The Monte Carlo simulation hence in use in the CDF B group contains $b\bar{b}$ events produced only via leading order processes and it does not reproduce $b\bar{b}$ events as they are seen in the CDF detector. Some studies [52, 53] show that, in order to predict the right $b\bar{b}$ production cross section at the Tevatron, additional processes have to

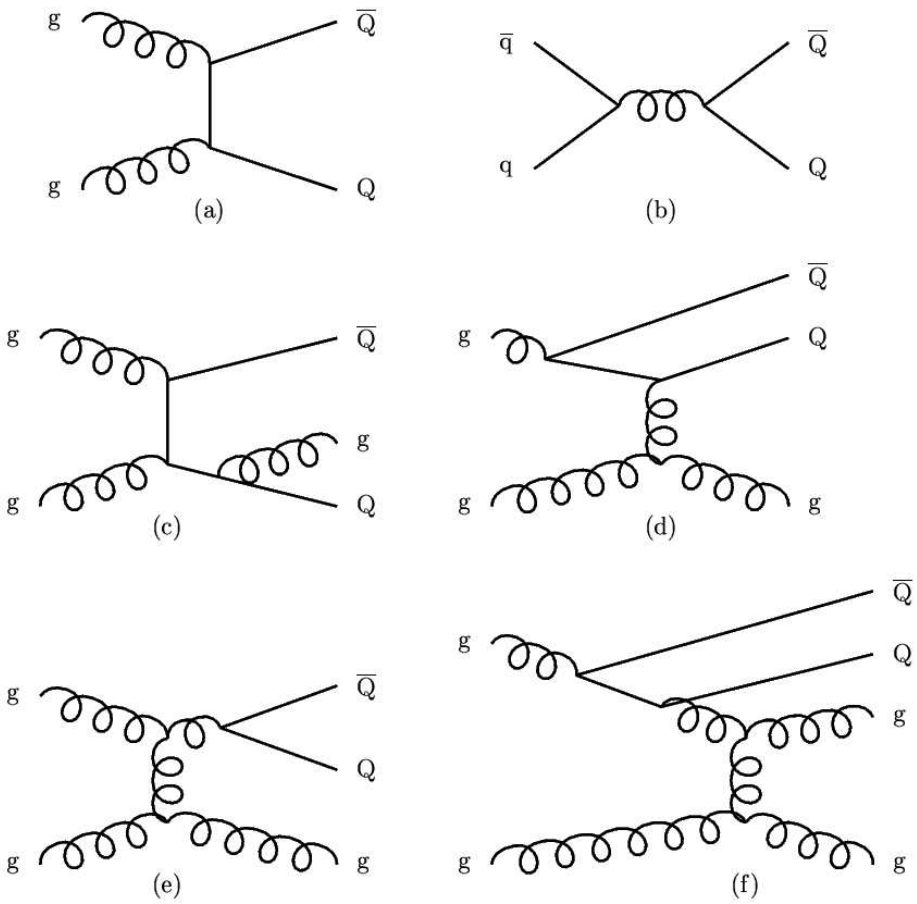

Figure 5.1: Feynman diagrams of the most important $b\bar{b}$ production processes. *(a)*, *(b)* and *(c)* show flavour creation processes. *(d)* is a flavour excitation diagram. *(e)* and *(f)* correspond to gluon splitting processes [52].

be added to a leading order Monte Carlo. The Feynman graphs associated to these production mechanisms are shown in fig. 5.1.
The leading order $b\bar{b}$ production process is *flavour creation* (FC). It creates the heavy quark pair through one of the processes illustrated by the diagrams in fig. 5.1a and 5.1b. The emission of a gluon by one of the heavy quarks (fig. 5.1c) does not affect the cross section of the process. The additional processes in the framework of the Parton Shower model are *flavour excitation* (FE) and *gluon splitting* (GS). Flavour excitation happens when a virtual heavy quark from the parton distribution of one incoming beam particle is put on mass shell by the momentum transferred through the interaction with a parton in another beam particle (fig. 5.1d). In gluon splitting processes the $b\bar{b}$ pair is produced in a gluon shower either by a final or initial state gluon (fig. 5.1e and 5.1f respectively). The above classification of $b\bar{b}$ production processes is

based on the number of heavy quarks in the final state of the hard scattering process (two for flavour creation, one for flavour excitation and none for gluon splitting). The three types of processes result in different correlations between the produced quarks. The study in ref. [52] shows that the $b\bar{b}$ production cross section by flavour excitation processes is comparable to the cross section of flavour creation at 1.96 TeV centre of mass energy. Gluon splitting plays a less important role but is nevertheless not negligible. As we shall see, all three of these $b\bar{b}$ production processes have to be included in the generation of a Monte Carlo sample to obtain a reasonable description of data.

5.2 Samples Description

A Monte Carlo sample containing $b\bar{b}$ flavour creation-only processes and a Monte Carlo sample containing additional $b\bar{b}$ generation processes have been produced. The generator used is Pythia [54], with the production of the additional processes has been switched off for the first sample and on for the second sample, and the ℓ+SVT trigger has been simulated for both samples. The b-hadrons have not been forced to decay in any particular channel, consequently the Monte Carlo samples include events in which the trigger lepton comes from a decay $b \to c \to \ell$. More details about the generation and simulation of the Monte Carlo sample are given in Appendix B. For sample with only flavour creation processes, the events have been simulated in the run range 138815 - 156487, corresponding to the data taken up to the shutdown in Jan. 2003 and the sample contains about 140k events. The simulated run range for sample with more processes is 138809 - 178785, which corresponds to the list of good runs taken up to Feb. 2004, and the number of events is about 113k.
In the follow up the second sample is often indicated as "all-processes" sample, or "sample with all processes", for conciseness.
The data sample used is a subset of the ℓ+SVT dataset[1] including runs until Feb. 04.

5.2.1 B Candidate Reconstruction

The ℓ+SVT trigger requirements are confirmed in the offline analysis[2]. The *signal B-candidate* is reconstructed in the same way for data and Monte Carlo samples by combining the lepton and the SVT track.
The background subtraction with the signed impact parameter of the SVT track δ_0^{SVT} is performed in the same way for data and Monte Carlo. The variable δ_0^{SVT} is computed

[1] Data have been processed with version 5.1 of the reconstruction software. The dataset name is `jbel0c` and `jbmu0c`.

[2] For this purpose the LeptonSvtSel module [55] is used.

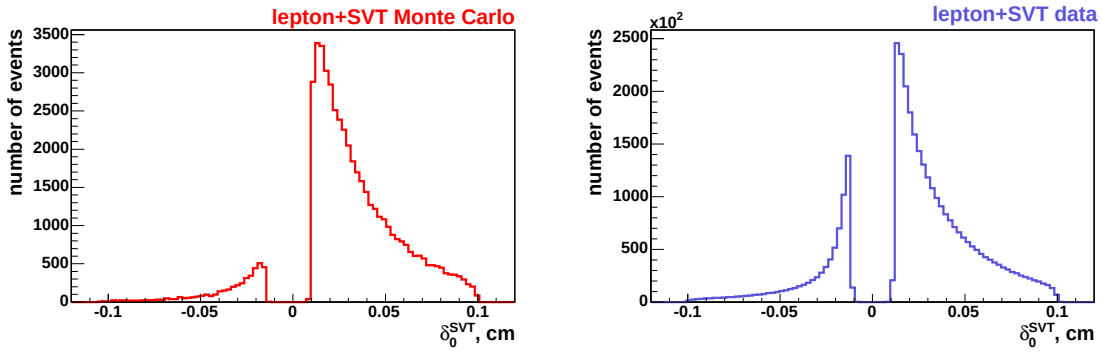

Figure 5.2: Distribution of the impact parameter of the SVT track signed with respect to the trigger lepton momentum δ_0^{SVT} for simulated events (*left*) and for data (*right*).

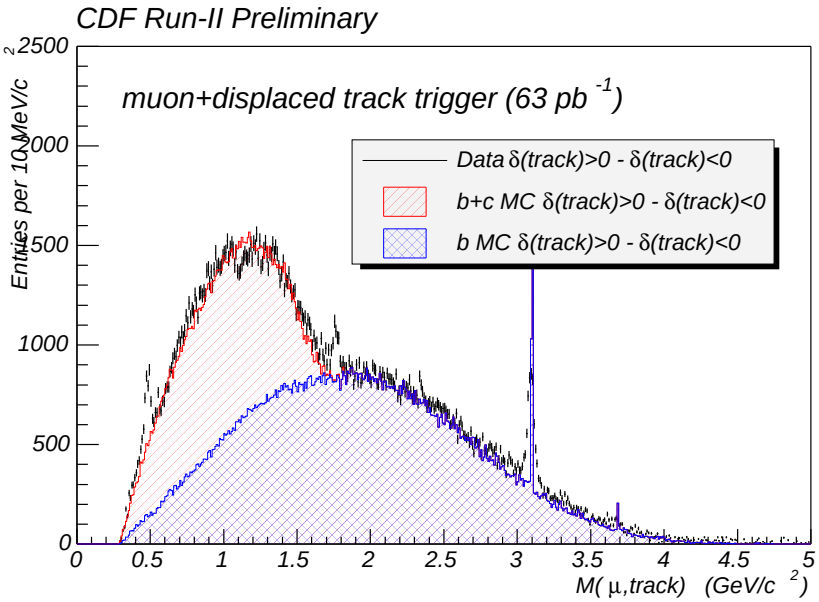

Figure 5.3: Distribution of the mass of muon and displaced track for data, $b\bar{b}$+$c\bar{c}$ simulated events $b\bar{b}$ only simulated events [56].

in every event as[3]

$$\delta_0^{SVT} = |\vec{d}_0^{SVT}| \cdot sign(\vec{d}_0^{SVT} \cdot \vec{p}_\ell) \tag{5.1}$$

The distribution of δ_0^{SVT} is shown in fig. 5.2 for Monte Carlo (left) and for data (right). The cuts at ± 0.01 cm and ± 0.1 cm correspond to the requirements at trigger level on the SVT track. For every variable the distribution obtained with the cut $\delta_0^{SVT} < 0$ is

[3]This definition is the same as in fig. 3.5 (left), where the track considered is the displaced trigger track and the trigger lepton momentum is substituted to the jet momentum $\vec{p}_{jet}$.

subtracted from the complementary distribution, i.e. obtained with the cut $\delta_0^{SVT} > 0$. This subtraction aims at rejecting the events in which the trigger lepton and the displaced track do not come from the same vertex, e.g. QCD background, for which the distribution of δ_0^{SVT} is symmetric about zero. About 42% of the events are removed in data. This fraction is only 17% for simulated events since QCD background is not included in the Monte Carlo sample.
It has been seen that requiring the mass of the B candidate to be the range [2 GeV/c^2, 4 GeV/c^2] suppresses the background from $c\bar{c}$ events [56]. The distribution obtained in that study for the mass of the B-candidate in the μ+SVT sample is shown in fig. 5.3. The candidates with mass larger than 4 GeV/c^2 are mostly background. The peaks corresponding to $K_S^0 \to \pi^+\pi^-$ ($\sim 0.5\,\mathrm{GeV/c^2}$) and $D^0 \to K^-\pi^+$ ($\sim 1.8\,\mathrm{GeV/c^2}$) are visible in data. In these decays one of the final state hadrons produces a fake trigger lepton. The $J/\psi \to \mu^+\mu^-$ peak is visible as well. The detector resolution has not been simulated in the Monte Carlo sample used for the study in [56], as is evident from the width of the J/ψ peak.
The procedure of cutting on the B-candidate mass and subtracting the background with δ_0^{SVT} has been approved by the CDF B group and is described in [56]. It is applied to every distribution and measurement shown in the following. The composition of the resulting sample is close to 100% $b\bar{b}$ events with negligible $c\bar{c}$ and light flavour component.

5.3 Comparison of Monte Carlo Samples

The estimates for the power of taggers at CDF Run II have been based on simulation containing only flavour creation processes. The result has proven to be too optimistic. It has become evident that additional processes cannot be neglected. In this section the differences between the samples and their comparison to data are illustrated. The complications added to the development of an opposite side tagger are discussed.
The study presented here is the first one within the CDF B group performed for tagging purposes on a Monte Carlo sample with additional $b\bar{b}$ production processes.

5.3.1 Quark Correlation

In each simulated event the non-excited b-hadrons are identified by accessing the Monte Carlo truth information. The b-hadron closest to the signal B-candidate direction is called *signal b-hadron* (or *true same side b-hadron*) and the other b-hadron in the event is the *tagging b-hadron* (or *true opposite side b-hadron*). The quarks that generated the b-hadrons are accordingly called *signal b-quark* and *tagging b-quark*.
The correlation between the signal and the tagging quarks is illustrated in fig. 5.4 and 5.5. Cuts at the generation level require that the signal quark is in the tracking

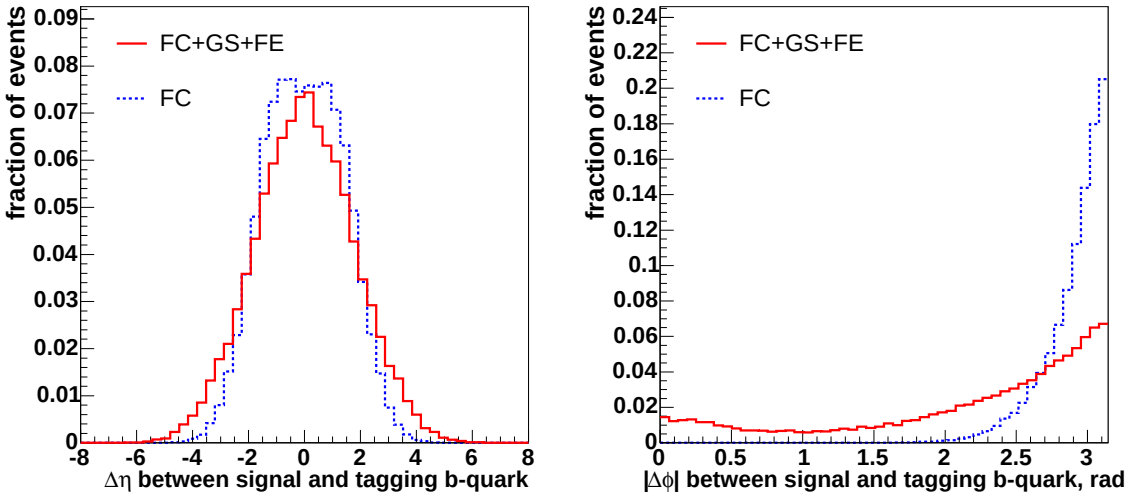

Figure 5.4: Correlation of signal and tagging b-quarks. *Left:* pseudorapidity difference between the quarks. *Right:* angular difference in the transverse plane.

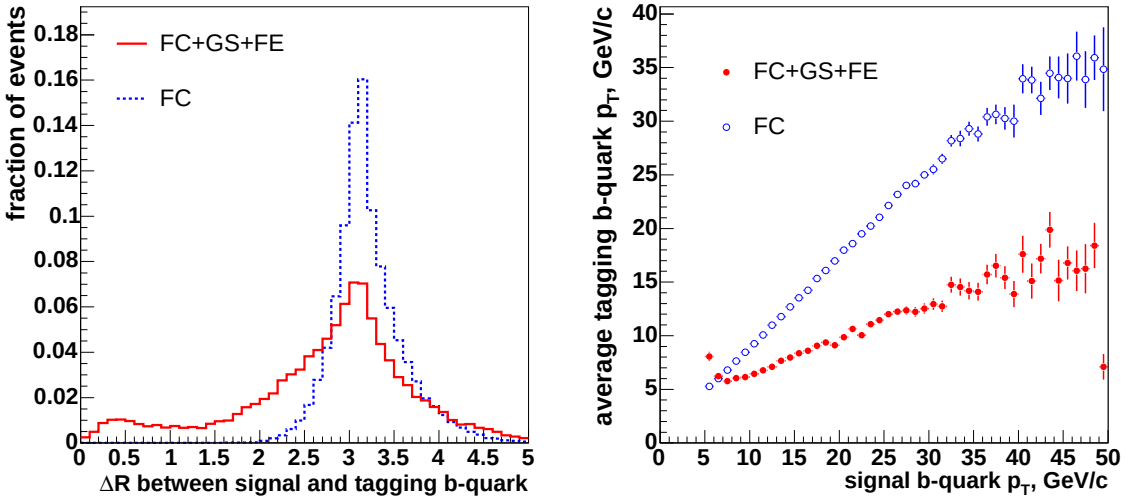

Figure 5.5: Correlation of signal and tagging b-quarks. *Left:* cone angle between the quarks *Right:* p_T correlation.

detector coverage ($|\eta| < 1.5$) and has a minimum p_T of 5 GeV/c. No cuts are applied on the tagging quark. The distributions on the difference of pseudorapidity ($\Delta\eta$, fig. 5.4, left), azimuthal angle ($\Delta\phi$, fig. 5.4, right) and cone angle (ΔR, fig. 5.5, left) are normalised to a unit area.

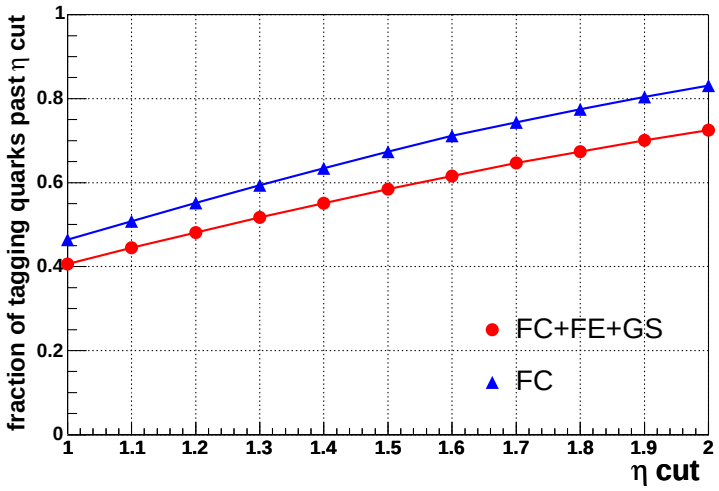

Figure 5.6: The signal b-quark is required to satisfy the cuts $p_T > 5$ and $|\eta| < 1.5$. The fraction of events in which the tagging b-quark $|\eta| < \eta_{cut}$ is shown. Different curves are drawn for the two Monte Carlo samples.

In flavour creation processes the b-quarks are produced preferentially back to back and the correlation between their momenta is strong. This does not hold for gluon splitting and flavour excitation. In the latter two cases the quark momenta are less correlated (fig. 5.5, right). In gluon splitting processes the quarks might be very close in space, thus rendering the separation of the event in same and opposite hemisphere impossible. In this case, it is difficult to properly flag a b-quark as signal and the other as tagging quark and wrong assignments can happen. The trigger p_T cut on the signal side biases the transverse momentum of the signal quark to higher values. In case of wrong assignments this bias is on the tagging side and this explains the rise at low momentum that one observes in the p_T correlation graph.
It is interesting to evaluate how often both b-quarks are in the detector coverage. The fraction of events for which the tagging quark has $|\eta| < \eta_{cut}$ is computed and shown in fig. 5.6 for different values of η_{cut}. The flavour creation Monte Carlo (triangles) overestimates the fraction by 6% to 11% with respect to the sample with all processes (circles). For $\eta_{cut} = 1.5$ the fraction is 58% for the all-processes sample and 67% for the flavour creation sample.

The comparison between the two Monte Carlo samples at generator level points out some differences that play an important role when developing opposite side taggers. According to the flavour creation Monte Carlo the heavy quarks are well separated in space and are both in the detector for most of the triggered events. Once the signal

Parameter	Value
minimum p_T for seed tracks $p_{T,min}^{seed}$	1.0 GeV/c
minimum p_T for non-seed tracks $p_{T,min}^{track}$	0.4 GeV/c
isolation cone around B candidate ΔR_{iso}	0.7
maximum cone to merge seeds ΔR_{clm}	1.5
maximum cone to merge non-seed tracks ΔR_{jm}	1.5

Table 5.1: Summary of the parameters for Cone Clustering optimised with respect to the tagging power of the cut base Jet Charge Tagger [47].

b-hadron direction is reconstructed in the transverse plane, the tagging b-hadron is likely to be found by looking in the opposite direction and its momentum is similar to the signal hadron momentum. This picture is over-optimistic and leads to wrong expectations for the performance of opposite side taggers.
The Monte Carlo with all processes depicts a more complex view. The tagging quark is more often out of the detector acceptance and sometimes very close in space to the signal quark. It is expected that the chance of defining the opposite hemisphere correctly and identifying the tagging hadron is consequently smaller in this sample.

5.3.2 Comparison to Data

In the following section we compare real data to Monte Carlo distributions of some opposite side related quantities. The opposite side is defined as the detector volume outside a cone of 0.7 around the reconstructed B candidate direction. The tracks on the opposite side have to satisfy requirements on the impact parameter ($|d_0| < 0.15$ cm) and on z_0 ($|z_0 - z_B| < 1$ cm, where $z_B = (z_\ell + z_{SVT})/2$), z_ℓ is the z-coordinate of the trigger lepton and z_{SVT} is the z-coordinated of the SVT track. These cuts reject tracks originating in primary interaction points different from the B production vertex [47]. Jets are reconstructed on the opposite side from the tracks remaining after the cuts. The jets are found by the Cone Clustering algorithm (see Section 3.4.3) with the tuned parameters used by the cut based Jet Charge Tagger (see Section 4.3.3 and ref. [47]), summarised in table 5.1. The jet selection of the cut based Jet Charge Tagger has been applied to choose a tagging jet among those reconstructed on the opposite side. All distributions shown in this section and in the following ones have been normalised to a unit area.
The number of tracks reconstructed on the opposite side is shown in fig. 5.7 for Monte Carlo and data. The distribution is better modelled in the Monte Carlo containing all processes than by the flavour creation only Monte Carlo. While the average number of opposite side tracks is very similar in the two Monte Carlo samples, the width of the distribution is larger for the all-processes Monte Carlo sample. This suggests that

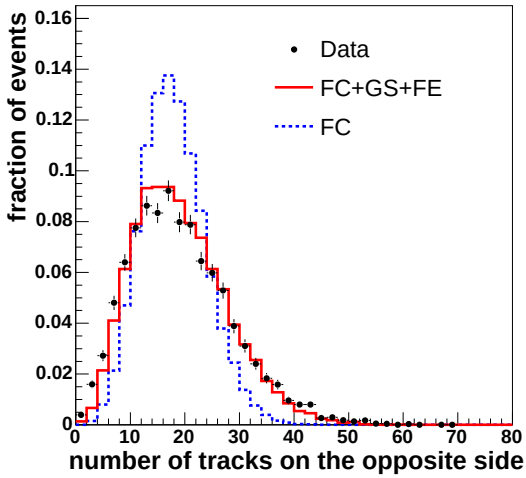

Figure 5.7: Number of tracks in the event outside a cone of $\Delta R = 0.7$ around the B candidate direction. The data points are compared to the distributions obtained for flavour creation only Monte Carlo (*dashed line*) and for all-processes Monte Carlo (*solid line*).

one of the two additional processes generates in average more particles than flavour creation and the other process generates fewer particles.
The average number of reconstructed jets on the opposite side (fig. 5.8, left) is smaller in the all-processes Monte Carlo than in the flavour creation only Monte Carlo. The overall distribution for the first sample is closer to the data. The distribution of the number of tracks in the jet tagged by the cut based Jet Charge Tagger (fig. 5.8, right) for data is more compatible with the all-processes Monte Carlo distribution than with the flavour creation only Monte Carlo. This is in agreement with the better description of the distribution of the track multiplicity on the opposite side given by the all-processes Monte Carlo. Finally, the distributions of $\Delta\phi$ between the tagging jet and the signal B candidate direction (fig. 5.9, left) and the cone angle ΔR between these directions (fig. 5.9, right) clearly show that the all-processes Monte Carlo reproduces the opposite side features of data events significantly better than the flavour creation only Monte Carlo.
A Monte Carlo sample containing only $b\bar{b}$ flavour creation processes does not correctly describe the data. Such a Monte Carlo sample might lead to a too optimistic estimate of the efficiency for the identification of the tagging b-hadron. Even without a fine tuning of Pythia fragmentation modelling, the all-processes Monte Carlo produced with default parameters describes data very well. In the following the flavour creation only sample will be abandoned and all studies will be performed with the sample containing all $b\bar{b}$ production processes.

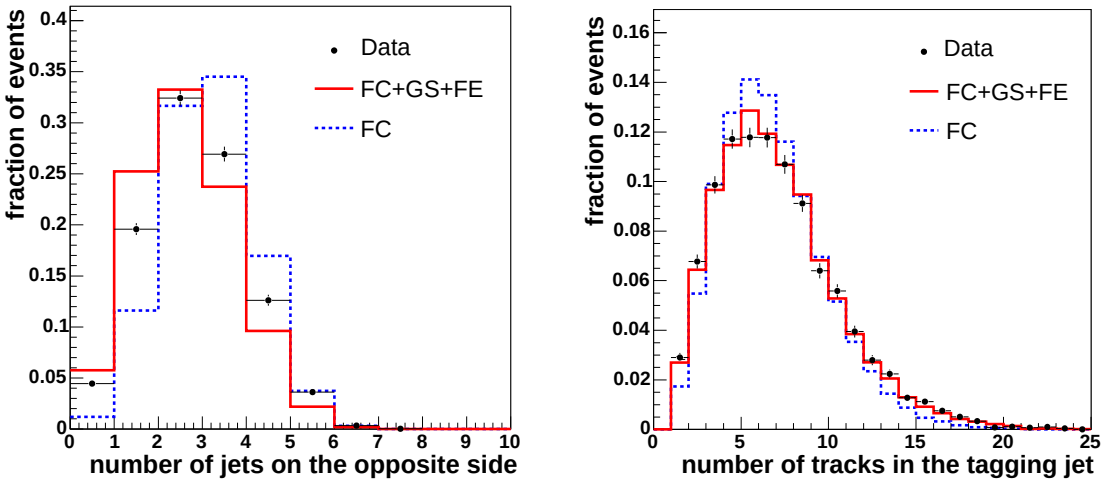

Figure 5.8: Comparison of distributions of opposite side variables for data (*dots*), flavour creation Monte Carlo (*dashed line*) and all-processes Monte Carlo (*solid line*). *Left:* Number of reconstructed jets on the opposite side by the cone clustering algorithm optimised as in [47]. *Right:* Number of tracks in the tagging jet as selected from the JetSelection algorithm, as used in [47].

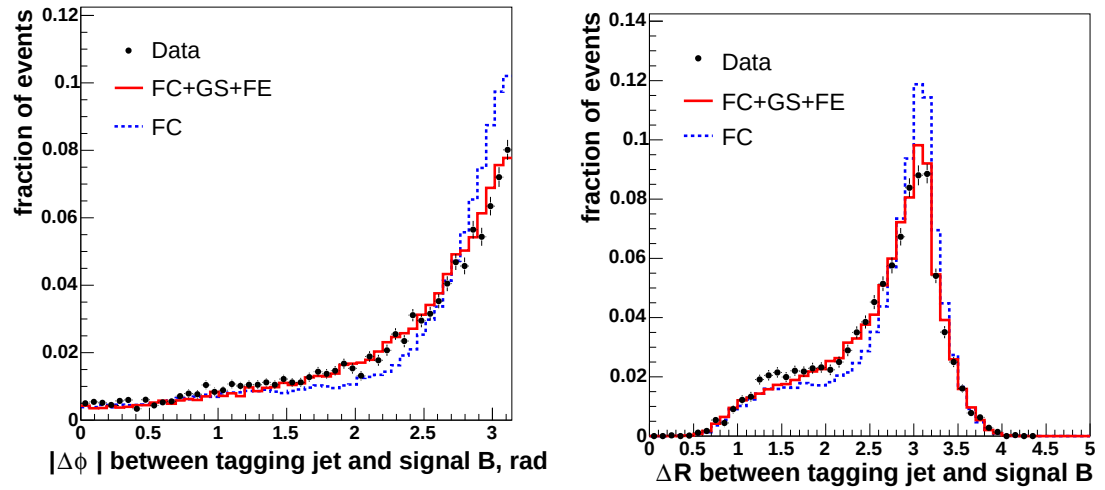

Figure 5.9: Comparison of distributions of opposite side variables for data (*dots*), flavour creation Monte Carlo (*dashed line*) and all-processes Monte Carlo (*solid line*). *Left:* angle in the transverse plane between the tagging jet direction and reconstructed signal B candidate. *Right:* cone angle ΔR between the tagging jet direction and the signal B candidate direction.

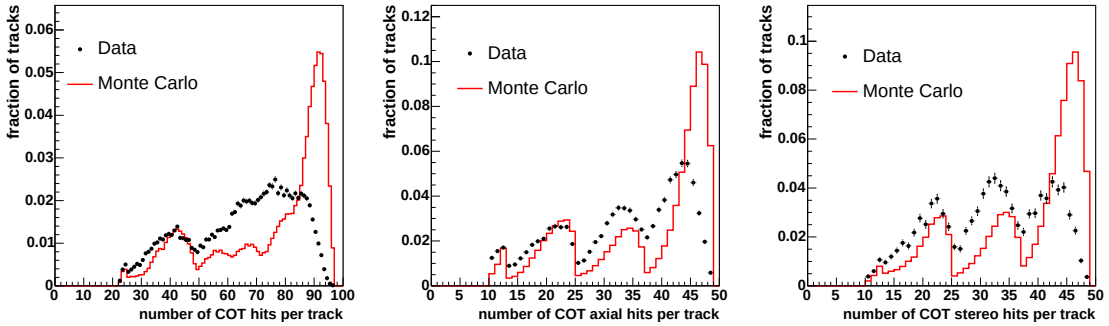

Figure 5.10: Distributions of the number of COT hits per track for data and Monte Carlo. *Left:* all hits. *Centre:* axial hits. *Right:* stereo hits.

5.4 Detector Description in Simulated Events

The accuracy of the detector description in the Monte Carlo sample is tested in this section. In particular some quantities related to the tracking detectors are compared. All tracks in the event with at least 10 stereo and 10 axial COT hits or with at least two r-ϕ hits in the SVXII detector have been re-fitted using the prescription of [57] for the energy loss correction for Monte Carlo and data.

The number of COT hits per track is shown in fig. 5.10. Obviously this quantity is not correctly modelled in Monte Carlo. The distribution for data is more smeared and the central value is smaller than in the simulation. Both axial and stereo hit numbers are different for data and Monte Carlo. The simulation of the COT is currently under development, so the discrepancy in the number of hits is expected to be reduced in future software releases.

The plots in the top row of fig. 5.11 show the comparison between data and simulation for the number of silicon hits per track. The Monte Carlo distributions are not in perfect agreement with data and tracks in simulated events appear to have on average more hits than in data. The impact parameter distributions corrected with respect to the beam spot (fig. 5.11, bottom left) present a small discrepancy in the tail. The error on the impact parameter corrected by the beam spot error (fig. 5.11, bottom centre) shows a disagreement as well. Moreover the correlation between impact parameter and its error is not perfectly modelled in the simulation: in the d_0 range of 0.02 cm - 0.1 cm the Monte Carlo resolution is better than the data resolution. Further investigation showed that the disagreement is mostly given by tracks that have their innermost ϕ hit in the layer closest to the beam line. In general the innermost hit and its error dominate the measurement of the impact parameter of the track and its error. The effect depends on the position of the hits of the track in the silicon detector and on the errors associated to the hits. The position of the hits in the silicon barrels and

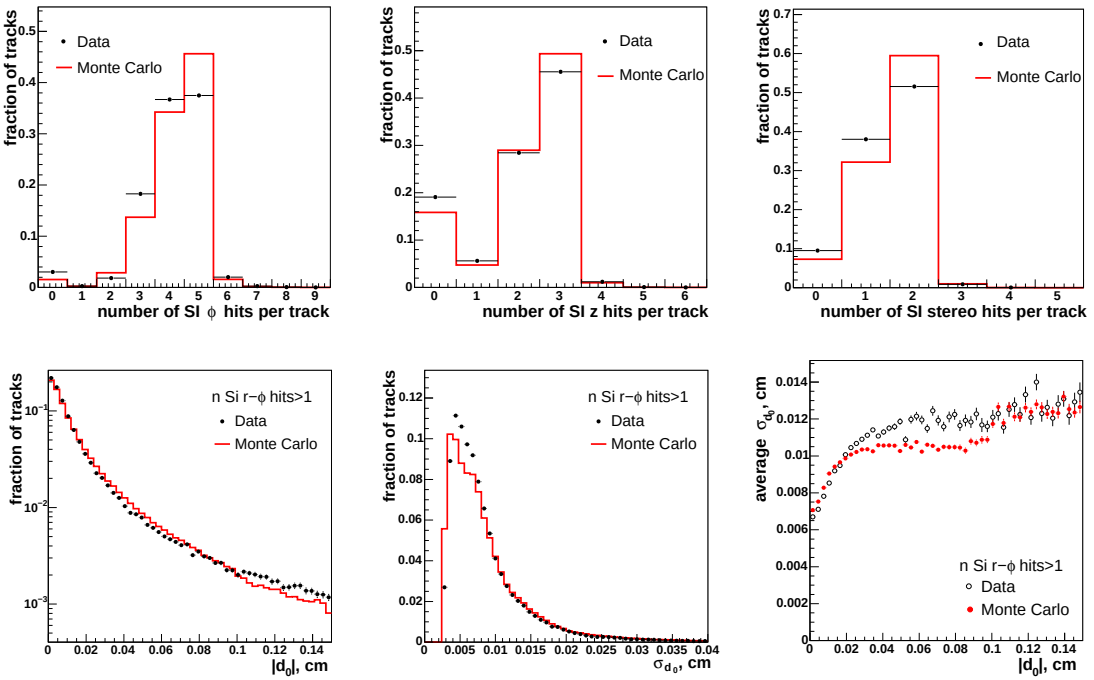

Figure 5.11: *Top row:* Distribution of the number of hits per track for data and Monte Carlo. *Bottom row:* Track impact parameter corrected with respect to the beam spot (*left*), error on impact parameter corrected by the beam line error (*center*) and their correlation (*right*).

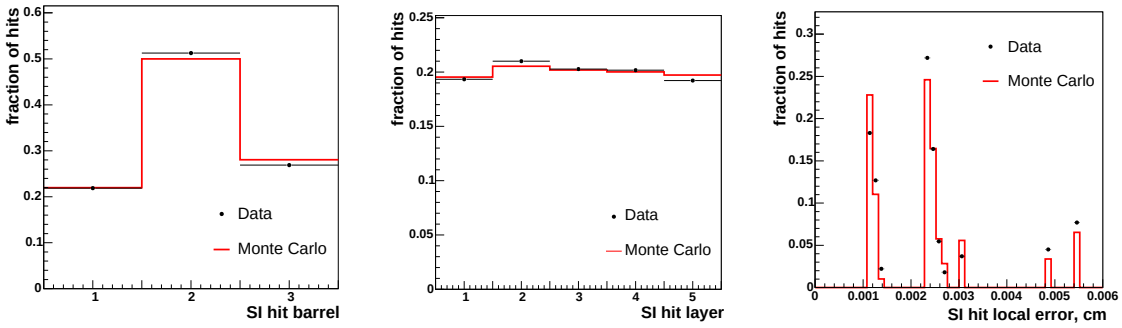

Figure 5.12: Hit position in the silicon detector barrels (*left*) and layers (*centre*), and hit error distributions for data and Monte Carlo (*right*).

in the silicon layers (fig. 5.12, left and centre) in the Monte Carlo is compatible with data. The error on the hit position (fig. 5.12, right) has a peak structure given by the resolution of the silicon strips in each layer. The substructure in the peaks is given by the number of strips that contribute to the hit. The error is in average bigger in

data than in simulated events. The disagreement is similar for each layer of the silicon detector. The Monte Carlo modelling of the hit error is wrong also in the innermost layer, which explains the disagreement seen in d_0 and σ_{d_0} distributions.

Since the hit content of tracks is not sufficiently well simulated, variables as the number of hits on tracks should not be used to parameterise other track properties, e.g. in a Neural Network, to avoid discrepancies in the results between data and simulation.

5.5 Event Shape Study

This section focuses on the kinematics of tracks on the same side, on the opposite side of the event and their relation between the sides of the event. The studies presented here are relevant to the development of opposite side taggers. The analysis of the same side distributions indicates how the tracks correlated to the signal b-quark could affect the tag of the b-flavour on the opposite side. The aim of the study of the opposite side is to understand the inclusive identification of the tagging b-hadron, e.g. with jet clustering algorithms. In the following "B" indicates any non excited b-hadron. This definition includes $B^0/\bar{B}^0$ ($\sim$40%), B^+/B^- ($\sim$40%), $B_s/\bar{B}_s$ ($\sim$10%) and $\Lambda_b/\bar{\Lambda}_b$ ($\sim$10%). "B^{**}"[4] will indicate any excited b-hadron. Plots regarding the distributions of the tracks corresponding to particles coming from the B decay, from the B^{**} decays and from the fragmentation of the b-quark into hadrons are shown in the following. These particles are identified with the Monte Carlo truth, excluding the particles originating from interactions with the detector. The B decay tracks include the daughters of intermediate-state particles originating from the B, such as D-mesons, J/ψ, etc. and the track categories are mutually exclusive. The average number of tracks of each type on the same and on the opposite side is shown in table 5.2. The number of reconstructed B decay tracks is on average lower on the opposite side because, unlike the signal b-quark, the tagging b-quark is not required to be in the tracking detector coverage. On both sides the fragmentation tracks are on average as numerous as the reconstructed B decay products. There are fewer tracks originating from B^{**} decay due to the small production rate for excited b-hadrons.

5.5.1 Same Side

As already mentioned, the signal B is reconstructed as the combination of the trigger lepton and the SVT track in each event. In order to understand the resolution on the reconstructed signal B direction, it is useful to look at the angle $\Delta\phi$ in the transverse

[4]This notation is used instead of "B^*" because B^* mesons decay dominantly into $B\gamma$, therefore no track from B^* is reconstructed.

track type	Same Side	Opposite Side
B decay	1.93	1.7
fragmentation	2.0	1.5
B^{**} decay	0.003	0.051

Table 5.2: Average number of tracks of each type on the same side and on the opposite side for simulated events.

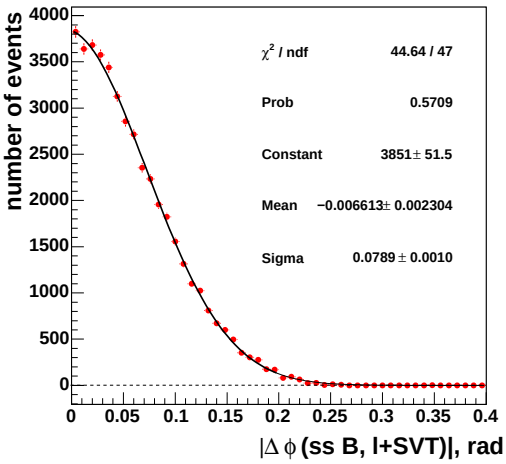

Figure 5.13: Angle in the r-ϕ plane between the true same side B and the B candidate direction. The solid line represents a gaussian fit to the distribution.

plane between the reconstructed B direction and the true direction (fig. 5.13). A gaussian fit of the angle distribution gives a width equal to 0.0789 ± 0.0010 rad, which can be interpreted as the resolution of the same side B direction in the transverse plane. Although the signal B is reconstructed from two tracks only, its direction in the transverse plane is determined rather precisely. The trigger tracks have relatively high momentum[5]. These are often the tracks carrying the highest fraction of the B momentum, so the sum of their momenta identifies the B direction well.

The properties of the tracks originating from the signal B decay are shown in fig. 5.14 and fig. 5.15. The transverse momentum distribution of tracks on the signal side (fig. 5.14, left) shows that the B decay tracks are more energetic than the fragmentation and B^{**} tracks. One notices also that the trigger cuts on momentum applied to the lepton and SVT tracks give a characteristic shape to the distribution of the p_T of the B decay tracks.

The tracks coming from the B decay and from the excited B decay are close to the

[5]The displaced track has a transverse momentum of at least 2 GeV/c and the lepton of at least 4 GeV/c (see Section 3.3.2).

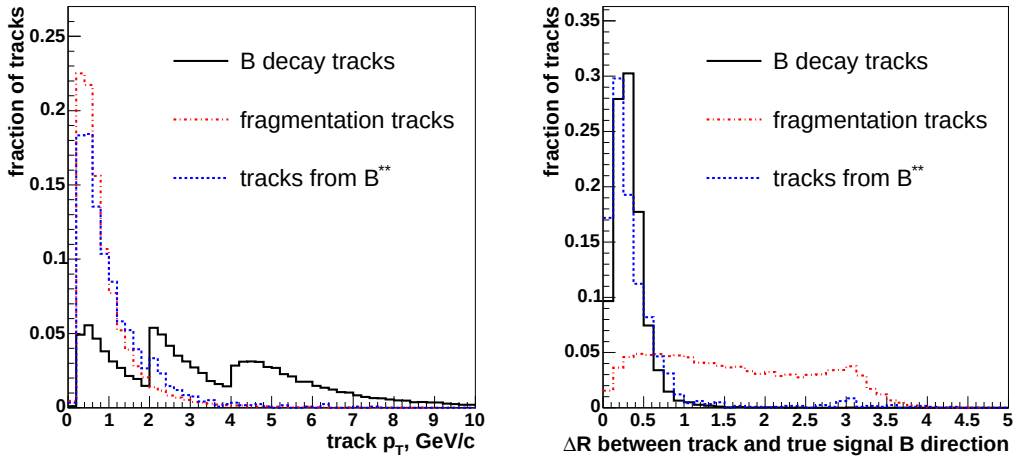

Figure 5.14: *Left:* Spectra of reconstructed transverse momentum spectra for tracks coming from B decays, from B^{**} decays and from fragmentation. *Right:* Cone angle of all tracks with respect to the true signal B direction, of tracks originating from the signal B.

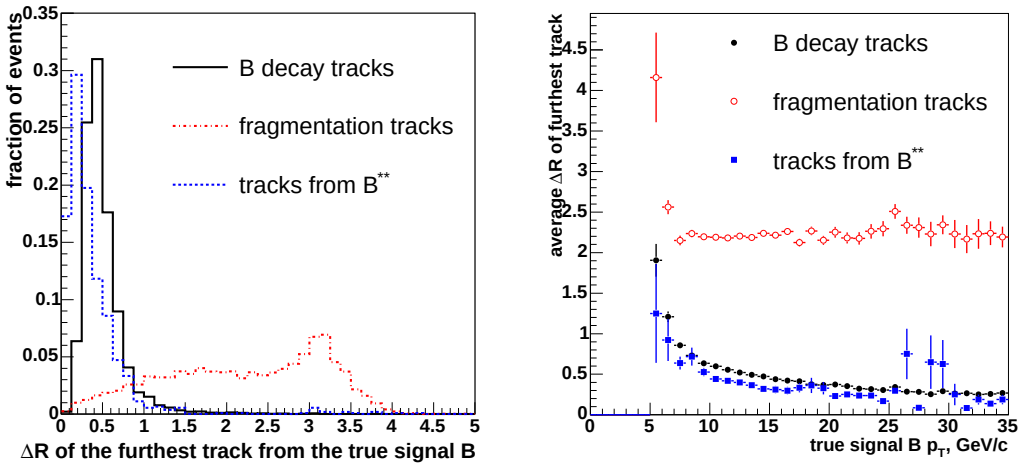

Figure 5.15: *Left:* Cone angle of the furthest track from the true signal B direction. *Right:* Dependency of the cone angle with respect to the true signal B direction of the furthest track on signal B transverse momentum.

B direction (fig 5.14, right). The plot in fig. 5.15 (left) illustrates the distribution of the cone angle of the track with the largest cone angle (*furthest track*) with respect to the true signal B direction . It shows that the tracks produced in the decay of B^{**} are closer to the B direction than those produced in the B decay and it indicates that the isolation cone of 0.7, already applied in the cut based Jet Charge Tagger,

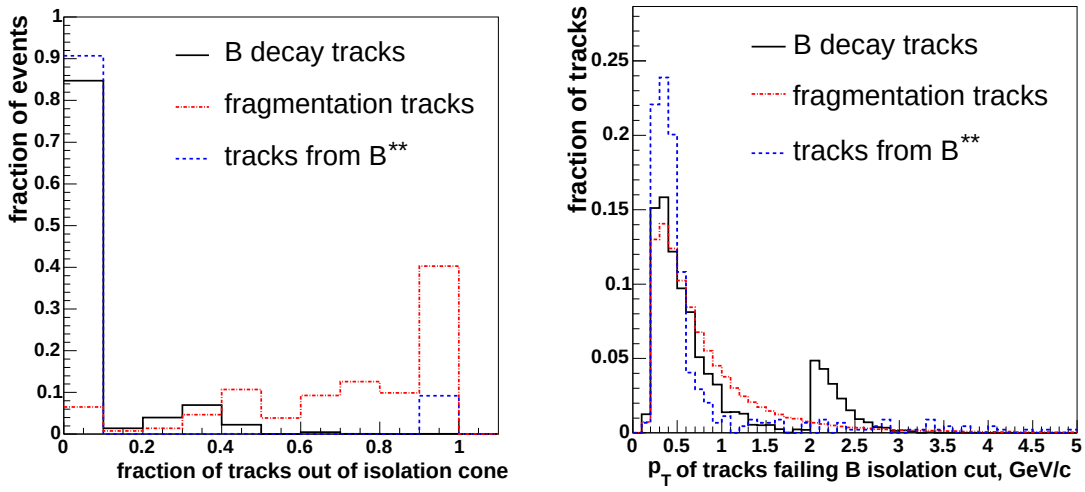

Figure 5.16: *Left:* Fraction of tracks related to the signal B that fail the isolation cut ($\Delta R = 0.7$). *Right:* Transverse momentum of the tracks failing the isolation cone.

rejects the tracks from signal B decay with high efficiency. The fragmentation tracks instead are be distributed uniformly in the detector. This behaviour does not depend on the transverse momentum of the signal B: the plot in fig. 5.15 (right) shows that for increasing transverse momentum of the true signal B, the B decay tracks become more collimated around the mother direction, but the average maximum distance of fragmentation tracks from the B direction does not vary significantly.
It is interesting to study the features of the signal and fragmentation tracks related to the same side B which are not included in the isolation cone, since these tracks are assigned to the opposite side and could affect flavour tagging. The fraction of tracks that fail the isolation cut is shown in fig. 5.16. In average, if particles from B or excited B decays are in the detector, $\sim$5% of their tracks fail the isolation cut. The average fraction of fragmentation tracks falling outside the isolation cone is $\sim$74%. It is not possible to completely isolate the fragmentation tracks and to define uncorrelated sides of the event. A big portion of fragmentation tracks originating on the signal side is always assigned to the opposite side. Since fragmentation tracks carry flavour information, their presence on the opposite side might affect the computation of the opposite side flavour. The p_T distribution of tracks failing the isolation cut is shown in fig. 5.16 (right). These tracks are in general soft, but about 20% of them has $p_T > 1$ GeV/c, which is the minimum p_T cut on seeds for the Cone Clustering algorithm (see Table 5.1), consequently they might seed the reconstruction of a jet on the opposite side.

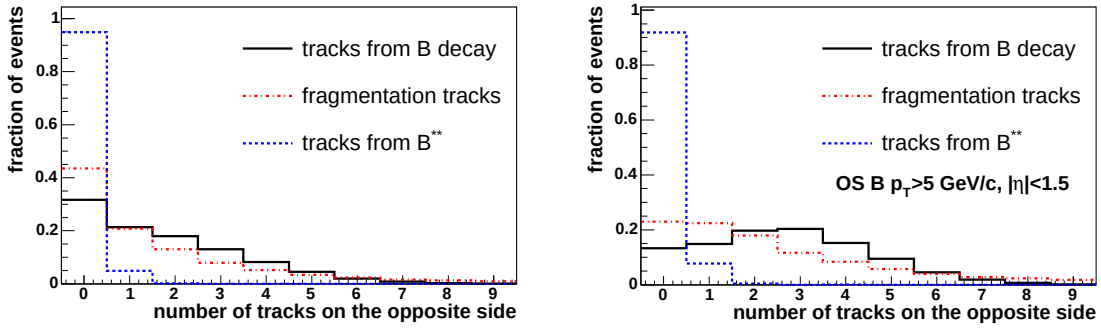

Figure 5.17: Distribution of the number of fragmentation tracks, B^{**} and B decay products reconstructed on the opposite side *Left:* no requirements on the opposite side b-quark *Right:* the opposite side b-quark is required to be in the detector acceptance and to have transverse momentum larger than 5 GeV/c.

5.5.2 Opposite Side

The distribution of the number of B decay products reconstructed on the opposite side is shown in fig. 5.17 in comparison with the number of fragmentation and B^{**} decay tracks. When no requirements are applied to the opposite side b-quark (fig. 5.17, left), no decay product from the tagging b-hadron can be detected in 32% of the events. When the tagging b-quark is in the tracking detector η-coverage and it has $p_T > 5$ GeV/c the fraction goes down to 13% (fig. 5.17, right). In these events, the direction of the momentum of the tagging b-quark is often very close the direction of the signal b-quark momentum, therefore its decay products fall inside the isolation cone. If in addition one requires an angular separation $\Delta R > 0.7$ between the momenta of the quarks, the fraction of events without decay tracks from the tagging b-hadron on the opposite side is 5%.
The true tagging b-quark does not have to satisfy any requirement at generator level in the Monte Carlo sample, in particular no transverse momentum cut is applied. Consequently the tagging B is on average less energetic than the signal B. The momentum distribution of its decay products reflects this (fig. 5.18, left). While the tagging B tracks are softer than the signal B tracks (compare to fig. 5.14, left) they all have a higher momentum than the fragmentation tracks on average, although 58% of them have p_T smaller than 1 GeV/c.
The tracks coming from the tagging B decay are less collimated than the tracks from the signal B decay. The furthest decay track might also be far away in angle from the true tagging B direction, as the long tail in the ΔR distribution in fig. 5.19 (left) indicates. A cone of 1.5 around the true tagging B direction collects most of the B decay tracks and excited B decay tracks, but it includes also a considerable fraction

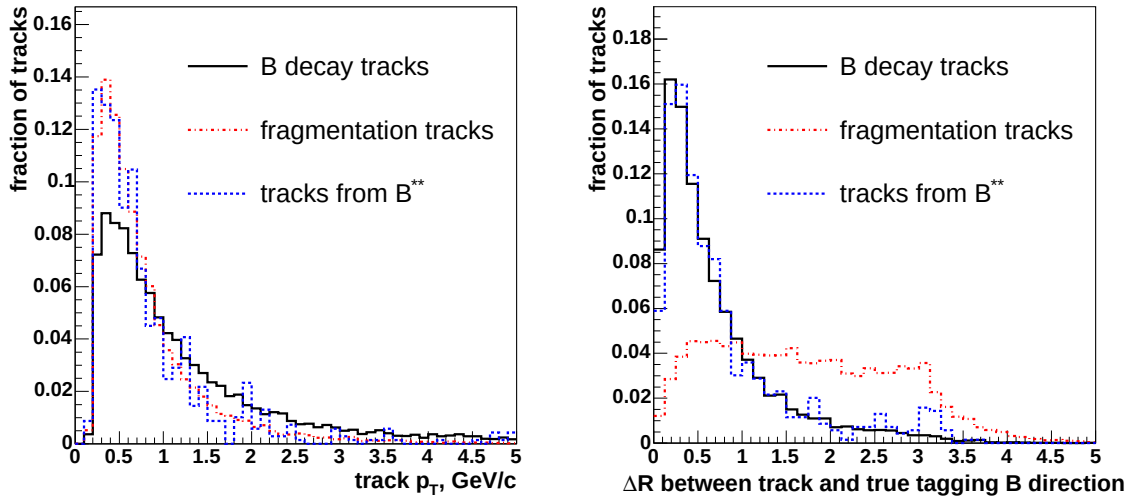

Figure 5.18: *Left:* Spectra of the reconstructed transverse momentum for tracks coming from B decay, from B^{**} decay and from fragmentation. *Right:* Cone angle distribution of tracks originating from the tagging B with respect to the true tagging B direction of all tracks.

of fragmentation tracks.
The tracks from B and excited B decays get closer to the true B direction as the heavy hadron transverse momentum grows (fig. 5.19, right). The maximum cone angle of the fragmentation tracks does not depend on the p_T of the B. Similarly to what is observed for the same side, it is not possible to find a cone around the B direction which could reject all fragmentation tracks.
In an ideal situation, as for example in the Monte Carlo with flavour creation only, the most probable tagging B direction could be found as the opposite to the direction of the signal B candidate. The tagging B transverse momentum could be estimated from the signal B p_T. A cone clustering algorithm could then use a cone size parameterised with the tagging B p_T. The parameterised cone size would allow the inclusion in the jet of most of the B decay tracks. It is clear that fragmentation tracks cannot be efficiently rejected by jet clustering. To reject fragmentation tracks impact parameter cuts, or similarly, track probability cuts are employed.
The reconstruction of pure jets with a cone algorithm is more complicated in reality. The Monte Carlo sample with all processes suggests that the signal and the tagging B directions are weakly correlated. Only in few cases the heavy hadrons fly back to back, consequently the signal B candidate cannot be used to estimate the tagging B candidate direction. The correlation between the transverse momenta of the heavy quarks is also weak, therefore the b-hadrons can have very different p_T. The tagging

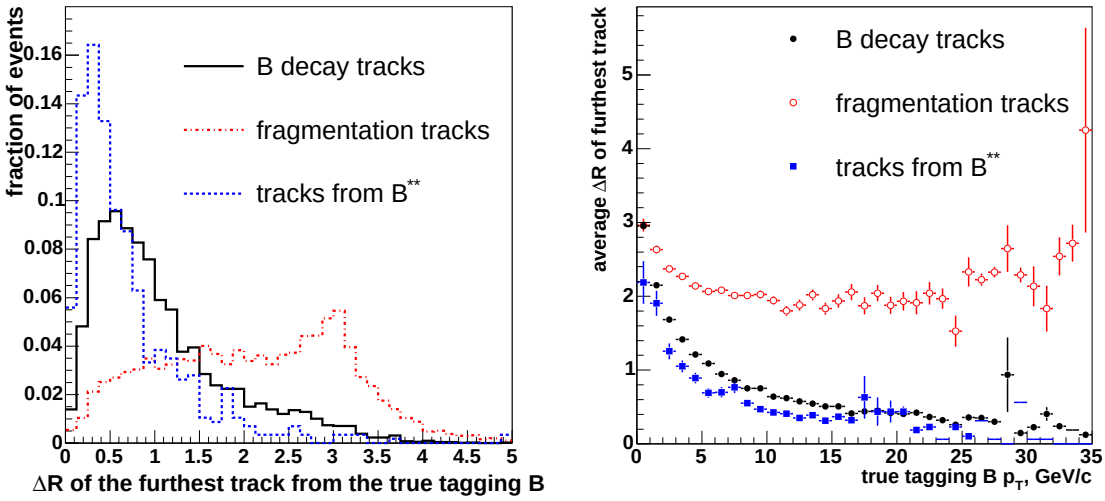

Figure 5.19: *Left:* Cone angle of the furthest track from the true tagging B direction. *Right:* Dependency of the cone angle with respect to the true tagging B direction of the furthest track on tagging B transverse momentum.

B p_T should then be estimated via an exclusive or an inclusive reconstruction method on the opposite side. The exclusive reconstruction though, would be successful only in a small fraction of events. The scenario suggested by the all-processes Monte Carlo is far from being ideal. The use of Cone Clustering to reconstruct jets containing most of the B decay products and the least of fragmentation tracks proves to be a challenging task.
A different approach in jet reconstruction is offered by Mass Clustering, described in Section 5.6.2. The invariant mass of all tracks on the opposite side is distributed as in fig 5.20 (left). The most probable value for the mass is rather high, around 15 GeV/c^2. The invariant mass of the tracks coming from the B decay is shown in fig. 5.20 (right). The mass is always smaller than 5 GeV/c^2 and it peaks evidently at the pion mass when there is only one charged particle from tagging B decay in the detector. The mass of a jet containing a very high fraction of B decay products should not have an invariant mass higher than the nominal B mass of $\sim$5 GeV/c^2.

5.6 Jet Clustering Algorithm Evaluation

In this Section the composition of jets reconstructed with Cone and Mass Clustering algorithms and with different jet sizes is studied. To judge the jet composition the *purity* and the *efficiency* are evaluated. The purity of a jet estimates the ability of the

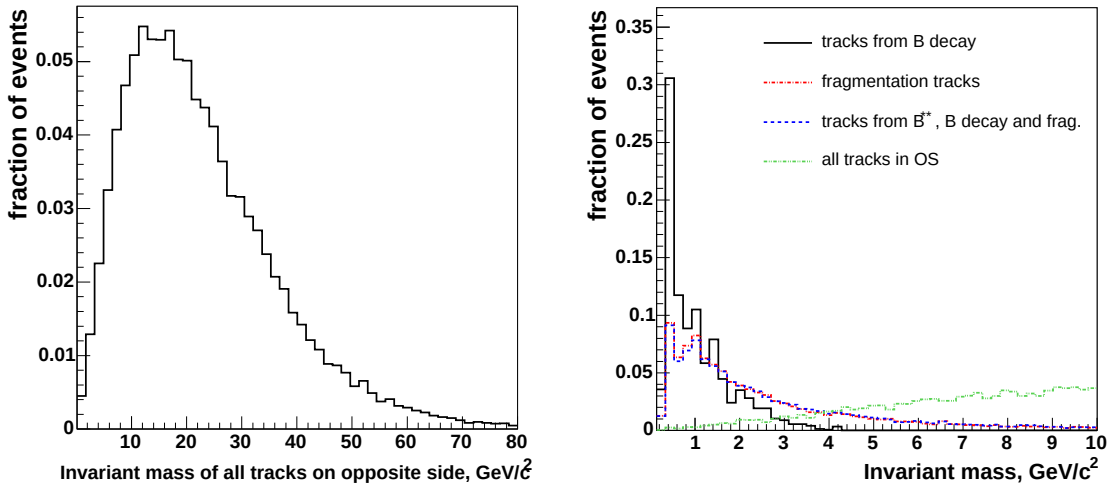

Figure 5.20: Invariant mass computed with the reconstructed momentum of tracks and with a pion particle hypothesis. *Left:* Invariant mass distribution of all tracks on the opposite side, including signal side fragmentation, B^{**} and B decays tracks that fail the isolation cone cut. *Right:* invariant mass of tracks coming from tagging B decay and/or from the fragmentation of the tagging b-quark, and comparison to the invariant mass of all tracks on the opposite side.

algorithm to include in a jet more B decay products (or fragmentation tracks) than other tracks and is defined as the number of tracks from B decay (or fragmentation) in the jet divided by the total number of tracks in the jet. The efficiency of a jet estimates how good the algorithm is in including the B decay products which are on the opposite side for a given event and is defined as the number of tracks from B decay (fragmentation tracks) in the jet divided by the number of B decay tracks (fragmentation tracks) on the opposite side.
The *signal jet* in the event is the reconstructed jet that contains the highest number of tracks from the tagging B decay. If there are two jets with numbers of B decay tracks that differ at most by one, the signal jet is the one closest to the true tagging b-hadron direction. In events in which all B decay products are outside the tracking detector, there is no signal jet. The efficiency and purity for a given algorithm are the average efficiency and the average purity of the signal jets that are reconstructed by the algorithm.

5.6.1 Cone Clustering

In order to understand which is an optimal value for the cone size used by the Cone Clustering algorithm and how the jet is composed, a scan of the cone angle parameter

Parameter	Value
minimum p_T for seed tracks p_T^{seed}	1 GeV/c
minimum p_T for non-seed tracks $p_{T,min}^{track}$	0.4 GeV/c
isolation cone around B candidate ΔR_{iso}	0.7

Table 5.3: Summary of the parameters used for Mass Clustering.

ΔR_{clm} is performed. The cone angle values used for this test are 0.4, 0.7, 1 and 1.5. The cone size for the merging of jets and non-seed tracks, ΔR_{jm}, is set equal to ΔR_{clm}. The other parameters are set to the optimised values listed in table 5.1. Plots of the performance of the Cone Clustering algorithm for different cone angle values can be seen in fig. 5.21 (open dots). As expected, the highest efficiency for collecting B decay tracks is given by the largest cone, the highest purity is given by the smallest cone (fig. 5.21, left). The reconstructed jets with $\Delta R = 0.4$ consist mostly of one track, so, although they are rather pure jets, they are not a good choice for tagging with Jet Charge. As the cone size grows larger, more fragmentation tracks are included in the jet (fig. 5.21, right). The growth of fragmentation track purity indicates that, by increasing the cone size, more fragmentation tracks constitute the jet, but not more tracks from B decay.
A hint on the tagging B direction resolution of the algorithm is given by the angle in the transverse plane between the best reconstructed jet and the true tagging B direction. A distribution of this quantity for each cone size is shown in fig. 5.22 (left). Only events in which the tagging B is in the detector acceptance ($|\eta| < 1.5$) and has a transverse momentum larger than 5 GeV/c are used to produce the residual plot. The distributions are very similar to each other, however the smallest width is obtained with the cone size $\Delta R_{clm} = 1$. Thus this cone size gives a better resolution on the tagging B direction. This conclusion agrees with the plot in fig. 5.18 (left), which indicates that in a cone $\Delta R_{clm} = 1$ around the true B direction there are ∼90% of the B decay tracks and ∼30% of the fragmentation tracks correlated with the opposite side b-hadron.

In the following jets will be reconstructed with a cone size of 1.5 in order to be able to use the SecVtx algorithm, which has been optimised on jets with $\Delta R_{clm} = 1.5$, within the Jet Charge Tagger.

5.6.2 Mass Clustering

A clustering algorithm that reconstructs jets on the basis of the invariant mass of pairs of tracks has been developed for the JADE experiment [58] and has been used at LEP.

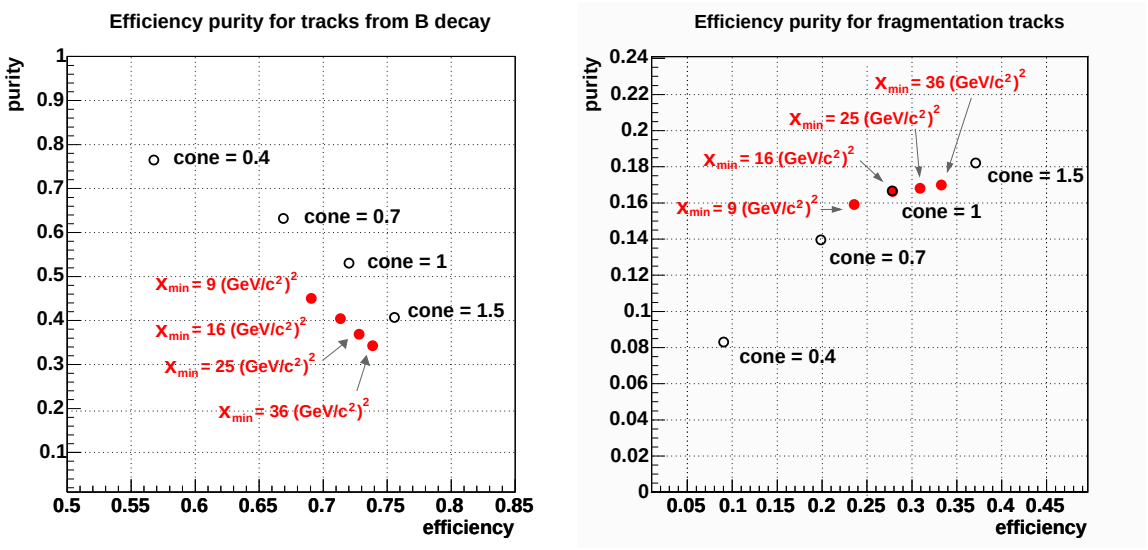

Figure 5.21: Evaluation of efficiency and purity for cone (*open points*) and mass algorithms (*full points*) for different cone values and x_{min} values.

CDF adopted it in Run I for the sin2β analysis [59].
The track pre-selection cuts are the same used for Cone Clustering (see Sections 5.3.2). The tracks seeding the algorithm have to have a transverse momentum larger than p_T^{seed}, which is a parameter of the algorithm. Clusters are constructed from each seed with each cluster corresponding to a single seed track, and the invariant mass of each pair of seeds is computed. Pairs that have invariant mass-squared m^2 smaller than a cutoff x_{min} are merged into a single cluster, starting with the pair with lowest mass. The merging procedure is repeated until no cluster pair has m^2 smaller than x_{min}. Non-seed tracks are added to the clusters with a similar procedure: m^2 is computed for every track-cluster pair, the tracks are included in the clusters if m^2 is smaller than x_{min}, starting with the lowest mass pairs. The merging stops when there is no track-cluster pair left with $m^2 < x_{min}$.
The result of the reconstruction of the jet with a cutoff on the minimum invariant mass is rather close to that of a cone based reconstruction. High p_T jets have a smaller cone size and lower p_T jets are broader. Figure 5.19 (left) suggests that this result is the ideal one. The algorithm could be tuned to give jets containing mostly B decay tracks. A simple scan of the invariant mass cutoff x_{min} is performed to evaluate the efficiency and purity of the Mass Clustering algorithm. The chosen values are 9 $(\mathrm{GeV/c^2})^2$, 16 $(\mathrm{GeV/c^2})^2$, 25 $(\mathrm{GeV/c^2})^2$ and 36 $(\mathrm{GeV/c^2})^2$. The minimum jet mass is the only parameter varied, all the other parameters are taken with their default values, listed in table 5.3.

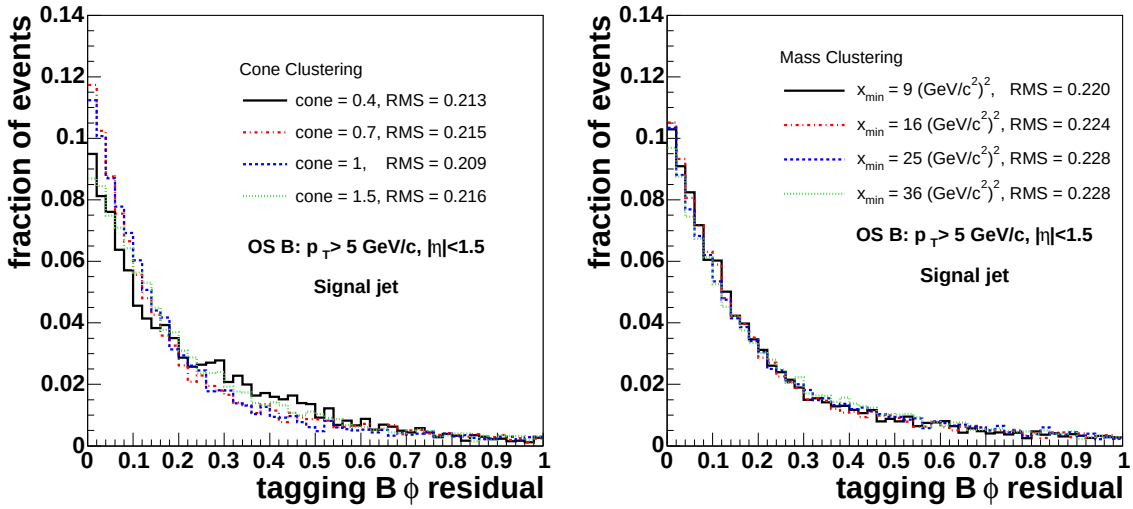

Figure 5.22: Angle in the transverse plane between the true tagging B direction and the signal jet reconstructed by the cone clustering algorithm (*left*) and by the mass clustering algorithm (*right*) in events in which the tagging B is in the detector acceptance.

The performance of the algorithm for different x_{min} values is shown in fig. 5.21 (full points). On average the mass algorithm is less pure than the cone algorithm. Less than half of the tracks in the signal jet are B decay tracks. Jets reconstructed with high x_{min} are as effective as jets with cone $\Delta R = 1.5$ in collecting B decay tracks. The fraction of fragmentation tracks included in the jet increases as the jet size becomes larger (fig. 5.21, right), although the purity and the efficiency vary only slightly.
The plot on the right in fig. 5.22 shows the angle in the transverse plane between the true tagging B direction and the signal jet reconstructed by the Mass Clustering algorithm. A cut is applied to select events in which the tagging B is in the acceptance and has a transverse momentum larger than 5 GeV/c. The distributions corresponding to different x_{min} cuts do not show significant differences, therefore x_{min} is not a parameter of primary importance in the tuning of the Mass Clustering algorithm. Other parameters, like the minimum transverse momentum of the seed tracks, might have a more significant effect on the resolution of the algorithm. The power of a Jet Charge Tagger with jets reconstructed by the Mass Clustering algorithm has been shown to be larger for lower values of p_T^{seed} [60]. The dependency of the jet purity on this parameter has not been studied yet.
It should be noticed that the resolution of the Mass Clustering algorithm is worse than the resolution obtained by the Cone Clustering algorithm using the cone size $\Delta R = 1$. Although the Mass Clustering algorithm has a stronger physical motivation

than the Cone Clustering algorithm, due to the cone size depending on the jet p_T, its performance with the default set of parameters is not better.
In Run I, jets reconstructed by Mass Clustering have been used in a Jet Charge Tagger for low p_T events [59]. Jets with fixed cone size have been utilised so far in Run II for opposite side tagging [47]. In the next chapters the possibility of employing jets reconstructed by both algorithms to identify the decay products and to tag the flavour of the opposite side b-hadron will be examined.

5.7 Summary

Before developing a b-flavour tagger for the opposite side, several features of $b\bar{b}$ events have to be understood. In this chapter many of these have been studied by means of simulated events. First of all it is legitimate to ask if the simulation traditionally employed for tagger studies, which includes only processes with two b-quarks taking part in the hard scattering (flavour creation), is adequate. A Monte Carlo sample comprising flavour creation and additional processes has been introduced and compared to data. The new sample reproduces the features of the data events better than the sample with flavour creation only processes. Moreover, the estimate of the power of an opposite side tagger is different in the two Monte Carlo samples, with the sample with flavour creation only processes giving a too optimistic estimation. Therefore tagging studies have to be performed on the Monte Carlo sample with the additional processes.
The questions relevant for a tagger and the answers found in this chapter are

- **How precise is the reconstruction of the signal B candidate?**
  The signal B-candidate is reconstructed inclusively as the combination of the trigger lepton and the displaced track. It has been found that this combination gives a satisfactory estimate of the true signal b-hadron direction, since a resolution of 0.0789 rad in the transverse plane was measured in the Monte Carlo sample.

- **How often is it possible to find both b-quarks in the detector acceptance?**
  The simulation indicated that the tagging b-quark is on the opposite side only for $\sim$60% of the events, which means that only for these events the definition of event sides is meaningful and the measurement of the b-flavour on the opposite side is possible.

- **Is the measurement of the b-flavour on the opposite side affected by the tracks that carry information about the flavour on the same side?**
  It has been found that the event sides are correlated: about 5% of the tracks

originating from the decay of the signal b-hadron falls out of the $\Delta R = 0.7$ cone and could be included in jets reconstructed on the opposite side. These tracks should be eliminated to avoid that they influence the measurement of the tagging power.
The fragmentation tracks are distributed isotropically in the event, rendering it impossible to isolate them with a cone cut. About 74% of fragmentation tracks correlated to the signal b-quark are found on the opposite side. They affect the measurement of the tagging power as well.

- **How many b-decay products can be found on the opposite side?**
  The opposite side contains on average $\sim$20 tracks. The average number of tracks originating from the decay of the tagging b-hadron is 1.7 and they are mostly low p_T tracks.

- **Is it possible to distinguish easily fragmentation tracks from b-decay products?**
  The fragmentation tracks are on average as numerous as the b-decay products on the opposite side and cannot be separated with a simple cone cut. Therefore a track probability variable, which exploits the displacement information, is going to be employed to distinguish B decay tracks from the products of the fragmentation.

- **Assuming that the opposite side b-quark reconstruction is performed inclusively with jets, is there a preference for the jet clustering algorithm to use?**
  Two algorithms have been considered: Cone Clustering, which builds jets with a fixed cone size, and Mass Clustering, which produces jets with cone size variable as a function of the jet momentum. Concerning the resolution on the opposite side b-hadron direction, no significant differences have been found between the two algorithms, although there is an indication that jets with a fixed cone size equal to one yield jets with higher resolution. The composition of the jets reconstructed by each algorithm has been examined. The Cone Clustering algorithm produces jets with on average fewer fragmentation tracks than the jets reconstructed by the Mass Clustering algorithm. The average number of B decay tracks included in jets is similar for the two algorithms. In the next chapters the usage of jets reconstructed by each algorithm is going to be investigated with respect to b-tagging and b-flavour tagging.

The results presented in this chapter show that the development of a powerful opposite side b-flavour tagger is a complex task. The problem is approached in the next chapter by combining the available information on b-decay products with a versatile tool that allows to exploit all available information in an optimal way: a Neural Network.

Chapter 6

b-jet Neural Network

The work presented in this chapter is based on the Monte Carlo studies described in Chapter 5. The main aim of this thesis is to develop, with the help of simulated events, a Jet Charge Tagger more powerful than the cut based algorithm [48] and with the ultimate goal of measuring the frequency of the B_s^0 oscillation. The strategy adopted is to provide a pure and efficient selection of the b-jet on the opposite side. At first a track probability variable based on Neural Networks is introduced to define a track-level b-tag. The track probability is exploited to develop a b-jet probability variable. The result is then applied to tag the opposite side b-quark flavour in the Jet Charge Tagger described in Chapter 7.
The track and jet probability presented here are a breakthrough since such an advanced analysis tool as a Neural Network had not yet been adopted by the CDF B group. Besides empowering the Jet Charge Tagger, the welcoming of Neural Network methods in the CDF B group opens the way to a combined tagger including particle ID information. This tool would effectively substitute all the opposite side taggers.

6.1 Sample Description

The data samples used for the studies presented here are the electron+SVT and the muon+SVT datasets[1]. The list of good runs recommended for analyses in the CDF B group has been used [61]. The sample includes data taken until August 2004 and does not contain the COT compromised and recovery run ranges, amounting to a total integrated luminosity of 355 pb^{-1}.
The Monte Carlo sample extensively used for this study has been described in Section 5.2 and B.2. It consists of $b\bar{b}$ events produced by leading-order and additional processes

[1]The name of the datasets are `xbel0d` and `xbmu0d`. Data have been processed with the offline software release 5.3.1

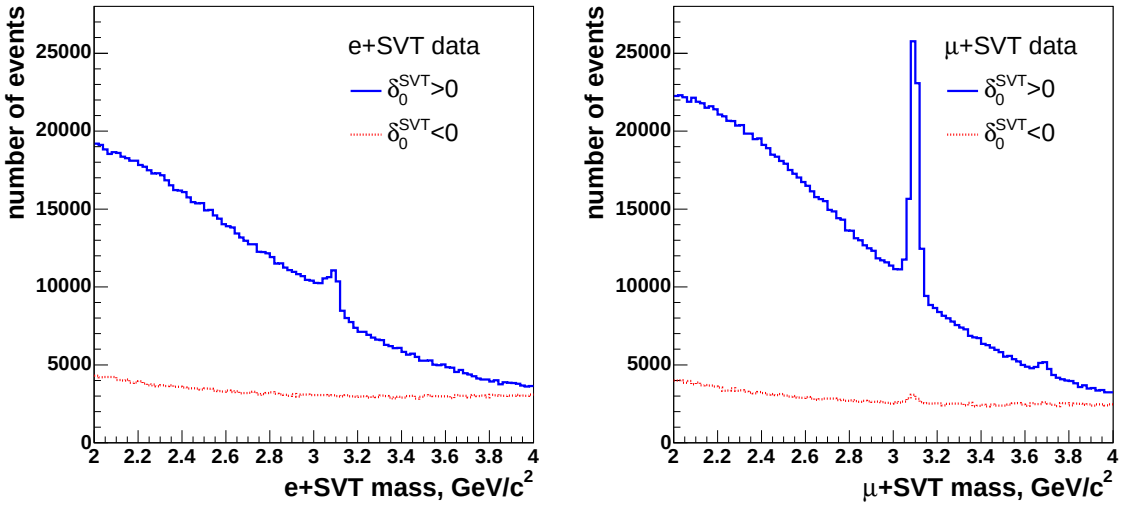

Figure 6.1: Invariant mass of the lepton and the SVT track for the electron (*left*) and the muon sample (*right*). The component corresponding to $\delta_0^{SVT} < 0$ (*dashed line*) has to be subtracted from the $\delta_0^{SVT} > 0$ component (*solid line*).

and fully simulated. The number of events in this sample is about 113k. The simulated run range corresponds to the list of good runs taken up to Feb. 2004.

6.1.1 Event Pre-Selection

The signal B candidate is reconstructed as the sum of the momenta of the trigger lepton and displaced track, as described in Section 5.2.1. Only one B candidate per event is considered, i.e. the one with the highest transverse momentum in case more than one is found. Similarly to Section 5.2.1, the candidates with mass outside of the interval [2 GeV/c^2, 4 GeV/c^2] are rejected and the background subtraction with the signed impact parameter of the displaced track (δ_0^{SVT}) is performed (see Section 5.2.1). The number of events remaining after the mass cut and the background subtraction amounts to $1.36{\cdot}10^6$ in data and $4.5{\cdot}10^4$ in the Monte Carlo sample.
The B candidate mass for the data sample is shown in fig. 6.1 separately for $\delta_0^{SVT} > 0$ and $\delta_0^{SVT} < 0$. The same distribution after the background subtraction is shown in fig. 6.2 in a comparison to the Monte Carlo distribution. The $J/\psi \to \ell^+\ell^-$ is visible at 3.1 GeV/c^2 and it is better resolved for the muon channel because of the smaller radiative tail. The peak at ∼3.7 GeV/c^2 corresponds to the decay $\psi(2S) \to \ell^+\ell^-$, not visible in the electron channel. The description of the mass spectrum in the Monte Carlo is satisfactory.

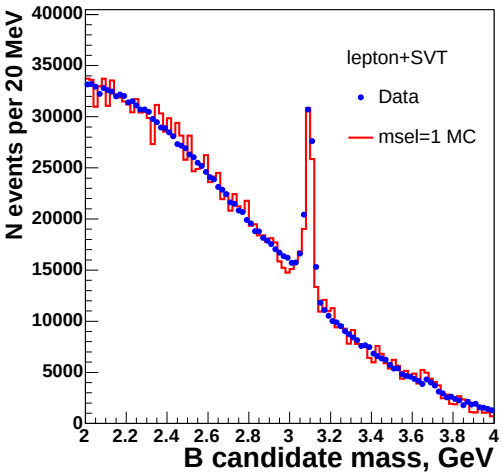

Figure 6.2: Invariant mass of the lepton and the SVT track for the Monte Carlo (*solid line*) and the ℓ+SVT data sample (*dots*) after background subtraction.

The tracks in the event are re-fitted and have L00 hits associated to them following the prescription of the CDF Tracking group [62]. A L00 hit is searched for only for tracks that have at least 3 hits in three different layers of SVXII to ensure a well defined search road for L00 hits[2].

6.1.2 Same Side B-Daughters Removal

Since the B candidate on the same side is partially reconstructed, the same side B daughters that leak out of the isolation cone might be included in the reconstructed jets on the opposite side (see Section 5.5.1). These tracks are correlated to the flavour of the same side B, thus they can affect the flavour measured on the opposite side. In the Monte Carlo sample 0.51% of the tracks on the opposite side come from the decay of the same side B. The fraction of same side B daughters in the sample is very small, nevertheless it should be further reduced.
Same side B daughters on the opposite side can be rejected by requiring that the invariant mass of one of these tracks with the same side B candidate is smaller than the B mass, about 5 GeV/c^2, and that their cone angle ΔR with the B candidate direction is small.
For each track on the opposite side the invariant mass with the lepton and the SVT track is computed using a pion hypothesis for each track. The invariant mass is

[2]L00 has been included very recently in the Jet Charge Tagger since it is expected to improve the tagging performance significantly. The development of the tagger has been carried on mostly without the use of L00.

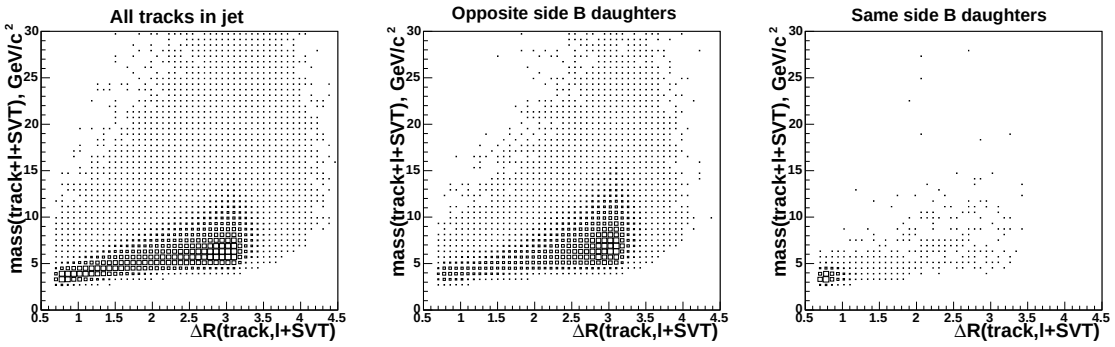

Figure 6.3: Invariant mass of tracks on the opposite side with the B candidate (lepton and SVT track) plotted versus the cone angle ΔR with the B candidate in simulated events. *Left:* all tracks on the opposite side *Centre:* Tracks originating from the opposite side B decay. *Right:* Same side B daughters.

shown in fig. 6.3 plotted versus the cone angle between the track and the B candidate. Monte Carlo truth matching identifies the tracks coming from the true b-hadrons in the event, as explained in Section 5.5. As clearly indicated in fig. 6.3 (right) the same side B daughters cluster at low invariant mass and low cone angle values.
The cases in which the invariant mass of same side B daughters is much higher than 5 GeV/c^2 are due to a fake lepton and/or a fake SVT track, i.e. events with the lepton and the SVT track coming from different vertices. These cases are mostly rejected with the background subtraction with δ_0^{SVT}. It is estimated that the fraction of events in which this happens in the Monte Carlo sample is 12.3% before the background subtraction. The number is reduced to 1.1% after subtracting the background. Indeed the number of tracks for which the invariant mass with the lepton and the SVT track is higher than 10 GeV/c^2 is negligible ($\sim$0.02%).
Some of the opposite side B daughters also populate the invariant mass/ΔR region where the same side B tracks concentrate. Most of the same side B daughters should be removed with a combined cut on the invariant mass and the cone angle without overly reducing the number of opposite side B products.
A mass/ΔR cut scan was performed on the Monte Carlo sample in order to find the combination that maximises $R_{SS} \cdot \epsilon_{OS}$ value, where R_{SS} is the fraction of same side B tracks rejected by the cut and ϵ_{OS} is the fraction of opposite side B tracks surviving the cut. The mass cut was varied in the interval [4.9 GeV/c^2, 5.8 GeV/c^2] and the ΔR cut in the interval [1, 1.9]. The optimal cut found corresponds to rejecting all tracks that have invariant mass smaller than 5.3 GeV/c^2 and ΔR smaller than 1.6. This combination rejects 92.8% of same side B daughters and keeps 92.2% opposite side B daughters. After the cut the fraction of tracks on the opposite side that are same side B daughters is 0.03%. The invariant mass and cone angle distributions for

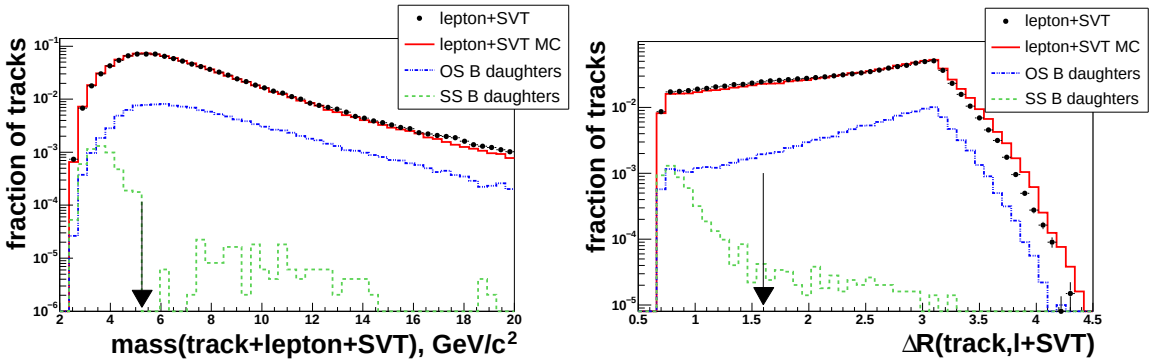

Figure 6.4: Comparison between data and Monte Carlo for the invariant mass of opposite side tracks with the B candidate (*left*) and for the cone angle between tracks and B candidate (*right*). The arrows point out the position of the optimised cut for the same side B daughters rejection (mass(track+lepton+SVT)<5.3 GeV/c^2 and $\Delta R < 1.6$)

Parameter	Value
cone definition for SecVtx tracks	1.5
minimum p_T cut for pass1 tracks	0.4 GeV/c
minimum p_T cut for pass2 tracks	0.4 GeV/c
maximum d_0/σ w.r.t. new vertex	3.0
d_0/σ for pass1 tracks	2.0
d_0/σ for pass2 tracks	1.0
minimum highest p_T for pass1 tracks	1.2 GeV/c
minimum highest p_T for pass2 tracks	2.0 GeV/c
minimum $\lvert L_{xy}/\sigma_{L_{xy}}\rvert$ for secondary vertex	3.0
move some pass1 tracks to pass2	false
use 3D vertex	false

Table 6.1: Summary of the optimised parameters for SecVtx [47]. The SecVtx algorithm has been introduced in Section 3.4.3.

data and Monte Carlo are shown in fig.6.4, where the arrows indicate the position of the optimised cut.

6.1.3 Jet Reconstruction

The tracks on the opposite side that pass the mass/ΔR cut have to pass additional cuts on the impact parameter ($|d_0| < 0.15$ cm) and on z_0 ($|z_0 - z_B| < 1$ cm), already mentioned in Section 5.3.2. Jets are reconstructed on the opposite side with the tracks remaining after the cuts.
The jets are found by the Cone Clustering algorithm with the parameters summarised

in table 5.1. The reconstructed jets are processed by the SecVtx algorithm (Section 3.4.3) to check for the presence of a displaced vertex. The parameters used for SecVtx-Alg are the same as those used by the cut based Jet Charge Tagger and are listed in table 6.1. Although in Section 5.6.1 it was shown that the jets with a cone size of 1.5 have a lower resolution than jets with cone equal to 1, the large cone size is kept due to the use of SecVtx. The steering parameters for SecVtx have been optimised for a cone size of 1.5 with the aim of maximising the power of the cut based Jet Charge Tagger. A re-optimisation in terms of jet purity could result in potential improvements although these are expected to be small. Therefore the parameters for the Cone Clustering algorithm are not re-optimised.

6.2 Neural Network based Track Probability

A high tagging dilution can be achieved by identifying the jet in the event that contains the highest possible number of opposite side B tracks.
The impact parameter (IP) based track probability described in Section 3.4.3 has been used by the cut based Jet Charge Tagger. Although the introduction of the IP probability has improved the tagging performance, there are some caveats related to its use. The probability has been calibrated on a high p_T data sample, with harder tracks than those in $b\bar{b}$ events in the ℓ+SVT dataset. The calibration has been performed using an earlier software release (4.11.X). Later releases brought reconstruction improvements from which also the IP probability could have profited. Currently the calibration is in the process of being updated with a later software release and the new version is not yet available. Finally, stringent hit requirements are applied to tracks before a meaningful probability value can be computed[3]. The opposite side tracks considered here do not have to pass any cut on the number of hits. It follows that not all the tracks assigned to jets can have an associated probability value.
To overcome these limitations a probability calibrated on a ℓ+SVT Monte Carlo sample with the latest reconstruction version and defined for every track associated to a jet is introduced here. This variable, instead of giving the probability that the track comes from the primary vertex, estimates how likely it is that the track is a B decay product. The signed impact parameter is combined with other information about the track and the jet to which the track is assigned. Jet variables are computed out of the tracks in the jet, so the available variables are correlated to some extent. A likelihood ratio does not offer the optimal combination because it does not take correlated inputs into account. A Neural Network can handle the correlations among the input variables and it is therefore potentially a more efficient tool.

[3]A track has to have at least 20 axial and 17 stereo hits in the COT and 3 to 5 r-ϕ hits in the SVXII detector.

6.2.1 Choice of Input Variables

The track probability Neural Network (*trackNet*, see Appendix C) is designed to tag whether a track comes from the B decay chain or not. At this point there is no distinction between same side and opposite side B tracks.
The *signal tracks* are defined to be all the B daughters, including the decay products of unstable B daughters. The *background tracks* are defined to be all the other tracks in the event. Each track is flagged on the basis of matching to the generator-level information.
With this definition of signal and background, the optimisation of the trackNet is performed on Monte Carlo. The result of the training is going to be applied to data, so it is important that the input variables and the correlation among them are correctly modelled by the simulation. The agreement of data and Monte Carlo is one of the criteria for selecting input variables in conjunction with the power to discriminate between whether the track was created at the primary or at a secondary vertex.
The set of input variables chosen for the trackNet is:

- track impact parameter with respect to the beam spot, $|d_0|$
- signed impact parameter significance $d_0^{signed}/\sigma_{d_0}$, where $d_0^{signed} = |d_0| \cdot sign(\vec{d_0} \cdot \vec{p}_{jet})$ (fig. 3.5)
- track transverse momentum p_T
- track was used for VxPrim vertex fit (Section 3.4.2). The flag is equal to 0 if the track was not used, 1 if the track was used, 2 if the primary vertex was not found in the event.
- ΔR of track with respect to the B candidate
- jet transverse momentum p_T^{jet}
- rapidity with respect to the jet axis $y = \frac{1}{2} \ln \frac{E+p_L}{E-p_L}$, where p_L is the longitudinal component of the track with respect to the jet axis, $p_L = \frac{\vec{p}_{track} \cdot \vec{p}_{jet}}{|\vec{p}_{jet}|}$

The impact parameter appears twice in the input list: in the magnitude of the track displacement, $|d_0|$, and in the significance of the displacement signed with respect to the jet. The first variable gives information only about the track; the second variable carries information about the track and the jet to which the track is assigned. The signed impact parameter should have a bigger discriminating power than the impact parameter alone, but the sign might be wrong due to an incorrect estimation of the B direction from the jet axis and to errors in the reconstruction of the track. In case of incorrect assignment the network learns that the signed d_0 significance is not useful

to characterise the signal, but the displacement information can still be used. The rapidity of the track with respect to the jet axis should be higher for signal tracks than for background tracks. The same trend is expected for the transverse momentum of the track.
The power to distinguish signal and background for some variables might depend on the event topology, for example on how back-to-back the b-jets are in the event, on the number of tracks in the event or on the B momentum. These variables are called *quality variables* because they do not themselves separate signal and background but provide the network with information on how reliable the discrimination of the other variables is likely to be. The ΔR and p_T^{jet} variables are quality variables for the trackNet because they express how jet-like the event structure is.
The resolution of $|d_0|$ and $d_0^{signed}/\sigma_{d_0}$ is influenced by the presence of a hit in the innermost silicon detector layer on the track. This effect is taken into account by optimising the trackNet separately for tracks with and without L00 hits. The category of tracks without L00 hits includes tracks with hits only in the COT, outside-in tracks, inside-out tracks and silicon standalone tracks (see Section 3.4.1) with less than 3 hits in 3 different SVXII layers and tracks that have at least 3 hits in different SVXII layers but did not have a L00 hit assigned. Figures 6.5 - 6.11 show the distributions of the input variables separately for tracks with/without L00 hits. The separation is meaningful only for the variables related to the impact parameter, the separated distributions for the other variables are shown for completeness. A comparison between data and Monte Carlo samples is shown as well. The data/Monte Carlo agreement is in general good.
The distribution of ΔR at small values is shaped by the same side B daughters rejection cut. The p_T distribution shows a peak at 1 GeV/c because the minimum p_T requirement for jet seeds enhances the sample of tracks in jets with tracks above 1 GeV/c.
Appendix C provides the details about the optimisation of the trackNet. The tracks with hits only in the COT have a worse impact parameter resolution than the tracks with hits in the silicon detector. For this reason, an alternative training of the trackNet has been performed on three subsamples, namely tracks with L00 hits, tracks without L00 hits but with at least a hit in the silicon detector and tracks with hits in the COT only. Since the tracks in the last category are very few in the sample, the training on the three subsamples has not proven to be necessary (the details are given in Appendix G.1). Therefore two subsamples have been used to achieve the results shown in the following.
Background tracks are expected to have output values close to 0 and signal tracks to have output values close to 1. The output can be interpreted as a probability that a given track is a B daughter (see appendix C.1). A distribution of the trackNet probability is shown in fig.6.12 for data and Monte Carlo. Tracks with a L00 hit have a

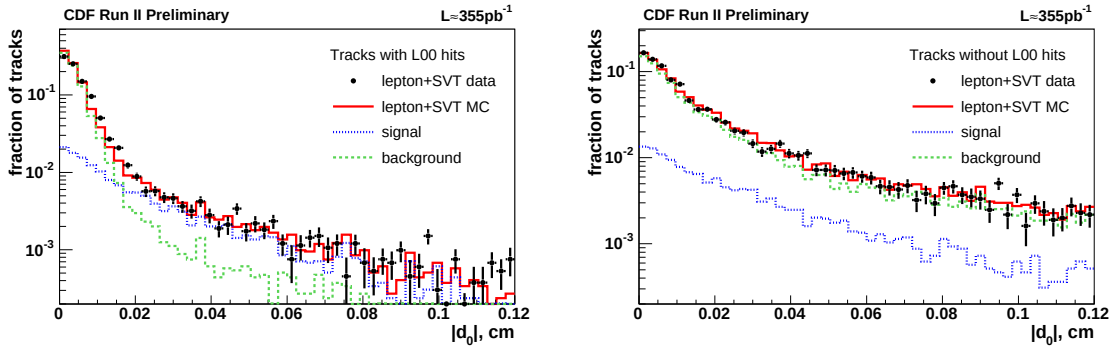

Figure 6.5: Distributions of impact parameter d_0. *Left*: Tracks with a L00 hit attached to them. *Right:* Tracks without L00 hits.

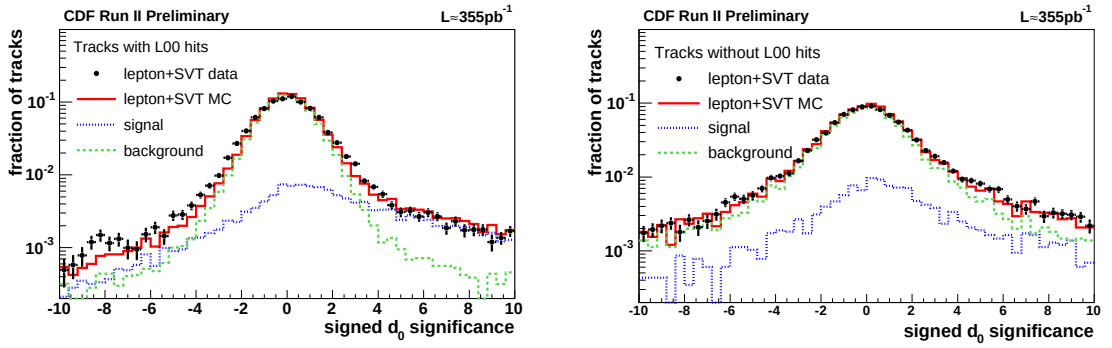

Figure 6.6: Distributions of the signed impact parameter significance $d_0^{signed}/\sigma_{d_0}$. *Left*: Tracks with a L00 hit attached to them. *Right:* Tracks without L00 hits.

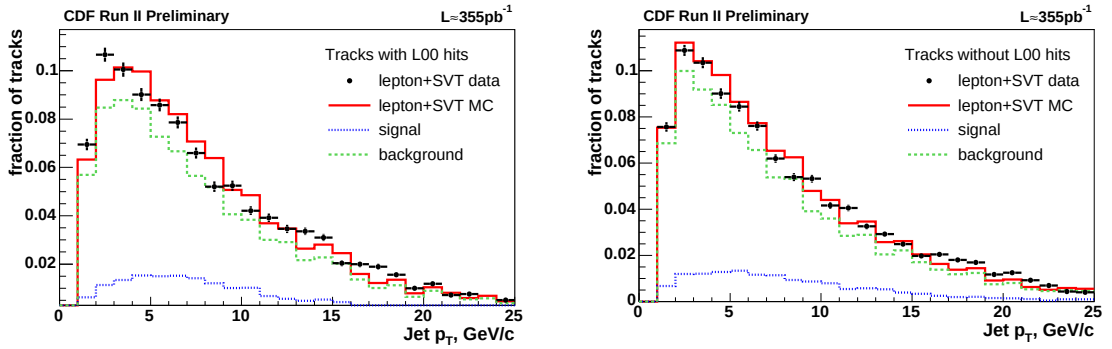

Figure 6.7: Distributions of jet transverse momentum. *Left*: Tracks with a L00 hit attached to them. *Right:* Tracks without L00 hits.

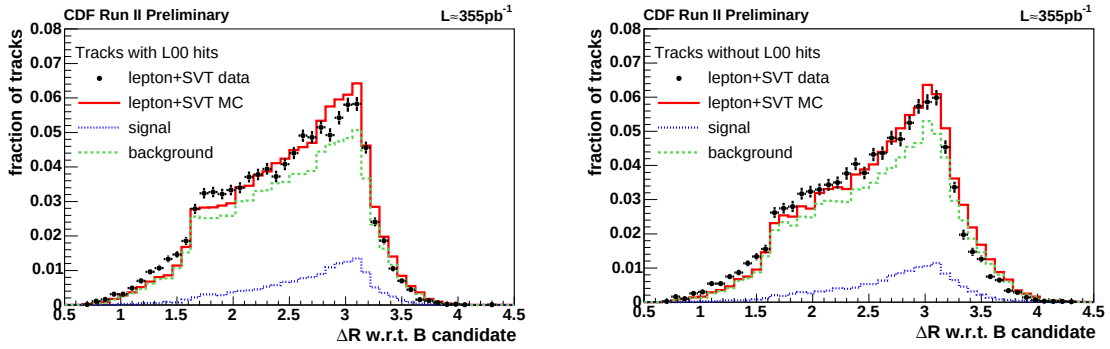

Figure 6.8: Distributions of the cone angle of tracks with respect to the B candidate ΔR. *Left*: Tracks with a L00 hit attached to them. *Right:* Tracks without L00 hits.

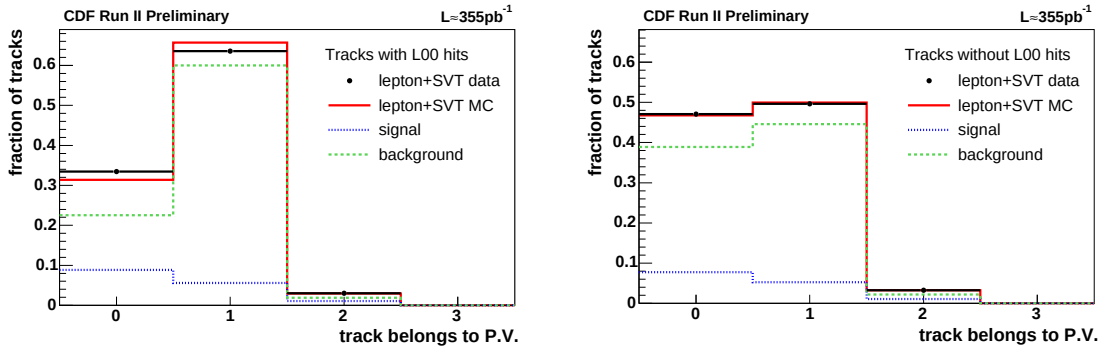

Figure 6.9: Distributions of the fraction of tracks used to fit VxPrim vertex. *Left*: Tracks with a L00 hit attached to them. *Right:* Tracks without L00 hits.

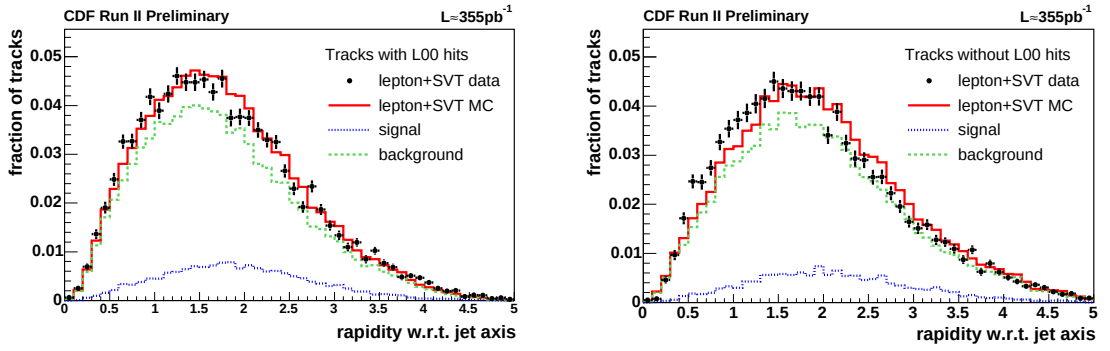

Figure 6.10: Distributions of the rapidity of the track with respect to the jet axis. *Left*: Tracks with a L00 hit attached to them. *Right:* Tracks without L00 hits.

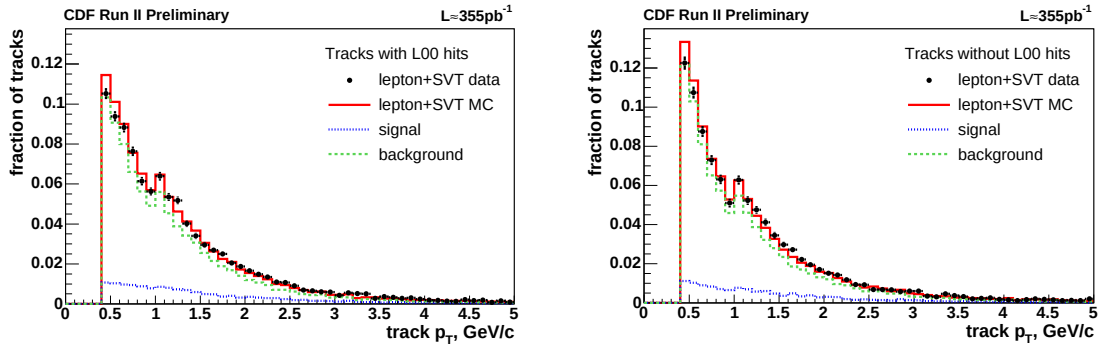

Figure 6.11: Distributions of the track p_T. *Left*: Tracks with a L00 hit attached to them. *Right:* Tracks without L00 hits.

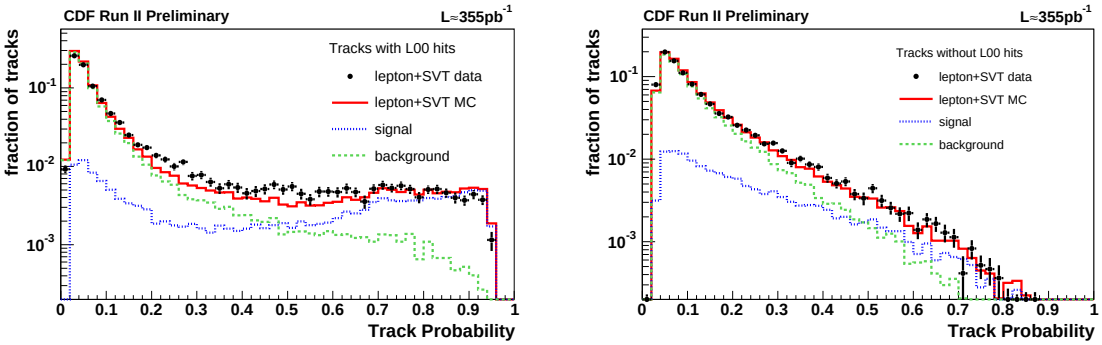

Figure 6.12: Track probability for the tracks with L00 hits (*left*) and without (*right*).

larger spread of trackNet outputs, due mainly to the higher resolution on d_0. The fact that the output for tracks without L00 hits is below 0.8 means that a purity higher than 80% in selecting these tracks cannot be reached because of the poorer resolution on the most powerful input variables, e.g. d_0.
It is important to notice that variables related to particle ID and to secondary vertices have been excluded from the input list to avoid correlations of the trackNet to other opposite side taggers. This choice was made so that a tagger could be developed that could be integrated into existing CDF B_s^0-oscillation analyses. The addition of particle ID would boost the performance of the trackNet and is advisable as a further development of the Neural Network based track probability.

6.2.2 Performance

The interpretation of the Neural Network output as a probability allows one to easily select a track sample with the preferred purity. To judge network performance, it is useful to form the efficiency for signal selection at a given cut on the trackNet output. Figure 6.13 shows the relation between purity of the sample and efficiency in selecting the signal for different cuts:

$$\text{purity} = \frac{N_{tracks}^{signal}(\text{output} > \text{cut})}{N_{tracks}^{signal}(\text{output} > \text{cut}) + N_{tracks}^{background}(\text{output} > \text{cut})} \tag{6.1}$$

$$\text{efficiency} = \frac{N_{tracks}^{signal}(\text{output} > \text{cut})}{N_{tracks}^{signal}} \tag{6.2}$$

Each point corresponds to a cut, and the cut increases when going from the right to the left side of the graph. The area under the curve gives an estimate of the trackNet performance. The performance of the trackNet for tracks with and without L00 hits is evaluated separately.
Figure 6.13 (right) compares the performance of the trackNet cut to that of the IP based track probability. It has already been mentioned that the IP track probability applies strict pre-selection cuts on the tracks and does not give a probability value for tracks that do not pass the pre-selection. The failing tracks are included in the denominator of the efficiency. The numerator of the efficiency includes all tracks with probability passing a given cut and having a meaningful value. Because of this the highest efficiency reached by the IP track probability is only about 45%.
Unlike the IP track probability, the trackNet is able to give a meaningful answer for each track in the jet. The trackNet output can be used to construct several jet variables that help in selecting b-jets. In this sense the trackNet is the building block of jet selection and is used for the jet Neural Network. In what follows, *track probability* denotes the trackNet output.

6.3 b-jet Selection Neural Network Variable

The potential of jets reconstructed on the opposite side of a $b\bar{b}$ event to be tagged as a b-jet can depend on many factors: the presence of the opposite side B in the detector acceptance, the momentum and the decay length of the opposite side B, the number of fragmentation tracks, etc. To approach b-jet selection in the most optimal way, the jet sample was split into three categories:

Class 1 Jets tagged by SecVtx algorithm (see Section 3.4.3) and with decay length significance of the secondary vertex $L_{xy}/\sigma_{L_{xy}}$ greater than 3

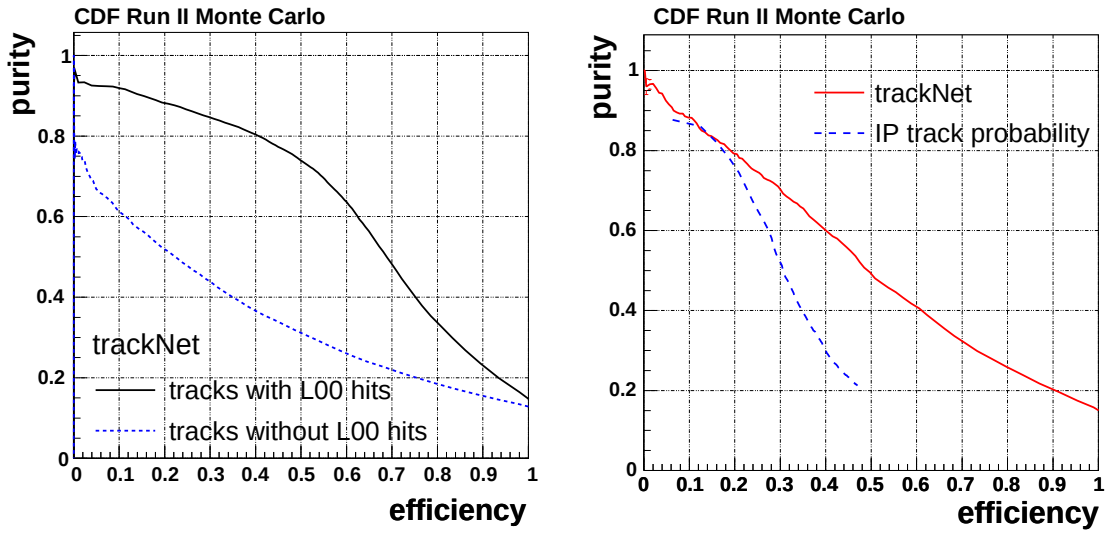

Figure 6.13: *Left:* Performance graph for the trackNet for tracks with and without L00 hits. Each point in the graph corresponds to a cut on the trackNet output. The performance is better for tracks with L00 hits thanks to the higher resolution on $|d_0|$ and $d_0^{signed}/\sigma_{d_0}$. *Right:* comparison between the performance of a cut on the trackNet output and a cut on the IP track probability. Tracks with and without L00 hits are not distinguished in this case, meaning that the solid line is equivalent to a weighted sum of the curves in the left graph.

Class 2 Jets not in Class 1 and with at least one track with probability greater than 50%

Class 3 Jets not in Class 1 and with no tracks with probability greater than 50%

The classes are mutually exclusive. The fraction of jets in Class 1 that do not have tracks with probability greater than 50%, and would then be assigned to Class 3 if they had not been tagged by SecVtx, is 22% in the data sample and 15% in simulated events.
Of all the jets reconstructed in the data and in the Monte Carlo samples, ∼5% of them belongs to Class 1, ∼15% belongs to Class 2 and ∼80% is in Class 3. The partition of jets into the three classes is charted in table 6.2.
A jet is a *signal jet* when it contains the highest number of tracks from the tagging B decay. In case two jets have numbers of B decay tracks that differ at most by one, the jet closest to the true tagging b-hadron direction is the signal one[4]. All the jets that are not signal are *background jets*. The fraction of signal jets in each class, estimated from the Monte Carlo sample, is also indicated in table 6.2. As expected, Class 1 is enriched in signal, thanks to the presence of an identified secondary vertex. The

[4]This definition of signal jet is the same given in Section 5.6

	Class 1	Class 2	Class 3	Total
e+SVT data				
N jets found	78000	243000	1269000	1590000
fraction	4.9%	15.3%	79.8%	100%
μ+SVT data				
N jets found	103000	330700	1714000	2147700
fraction	4.8%	15.4%	79.8%	100%
Monte Carlo				
N jets found	4450	11800	61700	77950
fraction	5.7%	15.1%	79.1%	100%
signal fraction	87.3%	65.2%	17.4%	28.6%

Table 6.2: Number of jets and fraction of jet type found for data and Monte Carlo. The purity for each subsample in the Monte Carlo is given.

sample of jets of Class 2 contains a similar amount of signal and background. Class 3 jets are the most numerous and are mostly background jets.
The classification of jets is a first rough b-jet tag which is refined by combining jet variables with a Neural Network. The advantage over a likelihood combination is the correct treatment of correlations. The knowledge of the jet class is not needed before computing the input variables because the training is performed for all jets with the same set of input variables. Although the training is not separate for the different jet classes, the information on which the jet classification is based is fed to the jet Neural Network: the secondary vertex tag (via the secondary vertex related variables) and the number of tracks with probability greater than 50%. The distribution of the jet Neural Network output is expected to depend on the jet quality.

6.3.1 Choice of Input Variables

Several variables that could distinguish signal and background have been considered as input for the b-jet Neural Network, *bJetNet*. Some of these are related to the secondary vertex properties, like the secondary vertex fit χ^2 probability, the momentum fraction and the number of tracks in the secondary vertex. Many variables are computed using the track probability information (weighted number of tracks, number of tracks with probability $> 50\%$, ...) and other variables are related to the kinematics and the shape of the jet (p_T, invariant mass, jet spread, ...). The bJetNet is trained on a variable that flags if the jet is signal or not.
The list of input variables for the bJetNet is:

- number of tracks with probability greater than 50%, N_{50}

- jet p_T
- Secondary vertex fit χ^2 probability, $P(\chi^2_{SV})$
- number of tracks in the secondary vertex, N_{SV}
- weighted number of tracks in jet $\sum_i t_i$, where the index i runs over all the tracks in the jet and t_i is the trackNet output for track i
- Jet probability J_P, where the jet probability is computed as in eqn. 3.2
- log of $J_P^{p_T}$, where $J_P^{p_T}$ is computed as in 3.2 and only tracks with $p_T > 1$ GeV/c are considered[5]
- highest probability for the tracks in jet, t^{max}
- Angular difference in the transverse plane between the jet and the B candidate, $\Delta\phi(jet, B)$
- transverse momentum of the track with the maximum momentum in the jet p_T^{max}
- $\sum_i p_{T,i}^{rel}$, where $p_{T,i}^{rel}$ is the magnitude of the track momentum component perpendicular to the jet axis
- $\sum_i p_{L,i}$, where $p_{L,i}$ is the projection of the track momentum along the jet axis
- track probability tag based on invariant mass, T_{nn}
- number of tracks, N_{tracks}
- jet invariant mass, m_{jet}
- jet spread, $\frac{\sum_i cos\alpha_i}{N_{tracks}}$, where $cos\alpha_i = \frac{\vec{p}_i \cdot \vec{p}_{jet}}{|\vec{p}_i| \cdot |\vec{p}_{jet}|}$

The invariant mass variable T_{nn} is computed as follows: The invariant mass of the tracks with the highest trackNet output is computed and the tracks are progressively summed up until the invariant mass becomes larger than the D^0 mass. The tag variable is then the trackNet output of the track that pushes the invariant mass value above the D^0 mass. b-jets are expected to have a certain number of displaced tracks that might come from the decay of a D meson. These tracks have on average high trackNet output and the D meson can be reconstructed inclusively by adding them together. It follows that the trackNet output for the last track added is large for jets

[5]Due to the minimum p_T requirement of 1 GeV/c for seed tracks, each jet has at least one of these tracks.

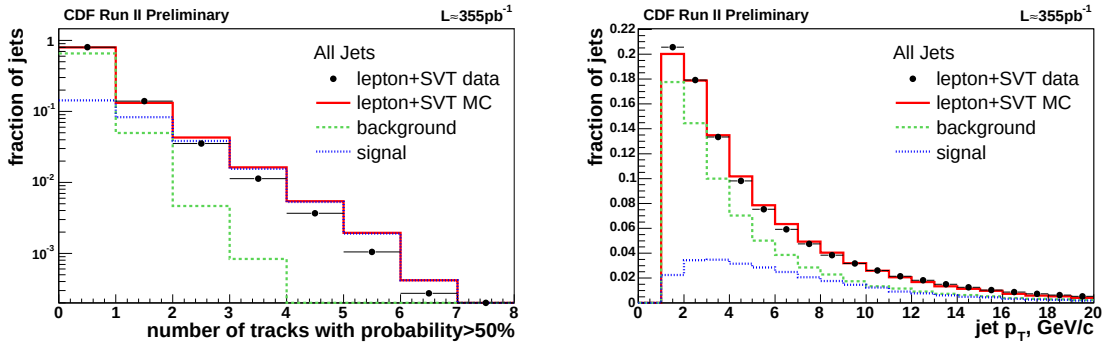

Figure 6.14: Input variables for the bJetNet: number of tracks with probability >50% (*left*) and jet transverse momentum (*right*).

including D mesons decay products.
The comparison of data and Monte Carlo for each variable is illustrated in figs. 6.14 - 6.21. The distributions are relative to all jets reconstructed in data and Monte Carlo events.
The input variables have been chosen at an earlier stage of the Tagger development, when L00 was not yet included. It was noticed that, after the addition of L00 hits to tracks, the data/Monte Carlo agreement became worse for some variables (highest trackNet output, weighted number of tracks, maximum track p_T). The distributions for one of these variables before the inclusion of L00 is shown in fig. 6.22 as a proof. The effect could be caused by a different efficiency in associating L00 hits to tracks in data and in Monte Carlo. As a consequence, the content of tracks with/without L00 is different for jets in data and simulation, leading to different contribution from the two track types in the computation of the jet variables. Since the inclusion of L00 improves the performance of the trackNet and bJetNet greatly, its use is still pursued. The power of the tagger based on the trackNet and bJetNet is later going to be measured on data, therefore the data/Monte Carlo agreement is of secondary importance at this point.

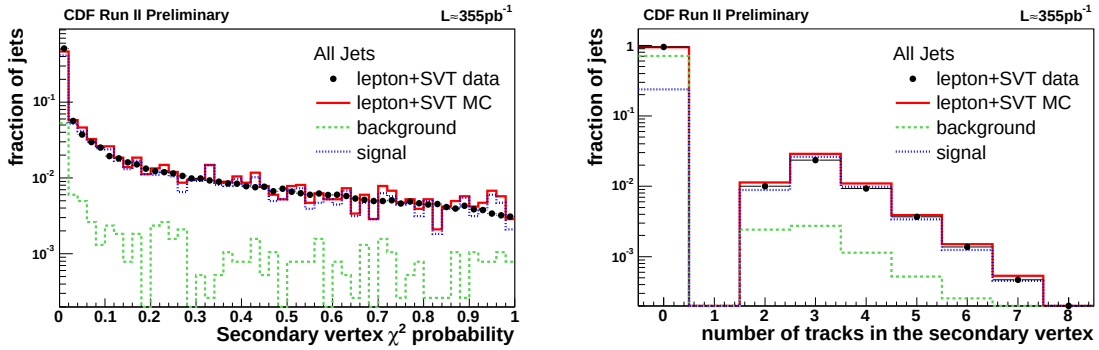

Figure 6.15: Input variables for the bJetNet: probability of the χ^2 of SecVtx fit (*left*) and number of tracks in the secondary vertex (*right*). For the first variable the fraction of background events is lower than the fraction of signal events because the χ^2 probability is defined only for Class 1 jets.

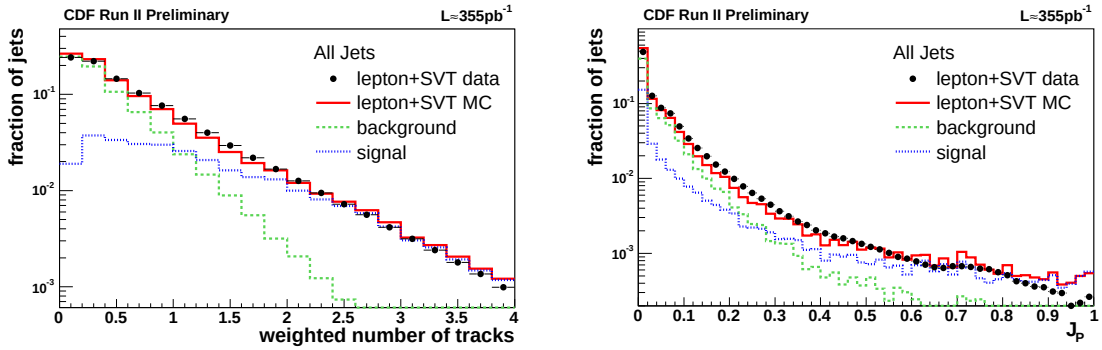

Figure 6.16: Input variables for the bJetNet: weighted number of tracks (*left*) and J_P (*right*).

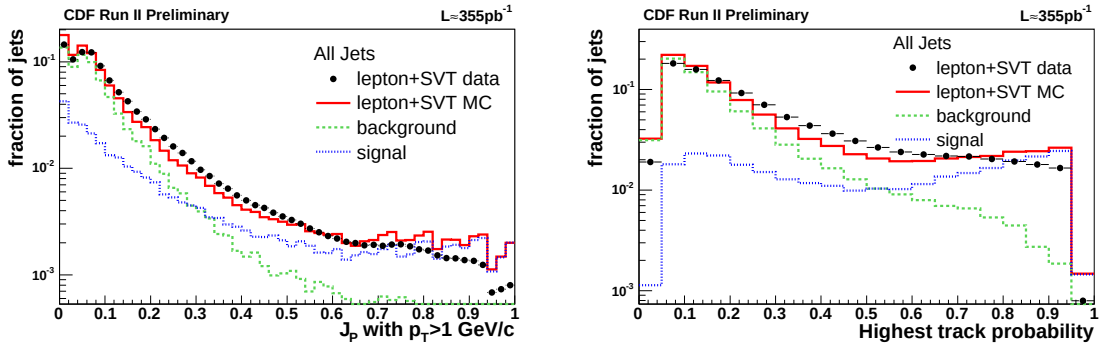

Figure 6.17: Input variables for the bJetNet: $J_P^{p_T}$ (*left*) and highest track probability (*right*).

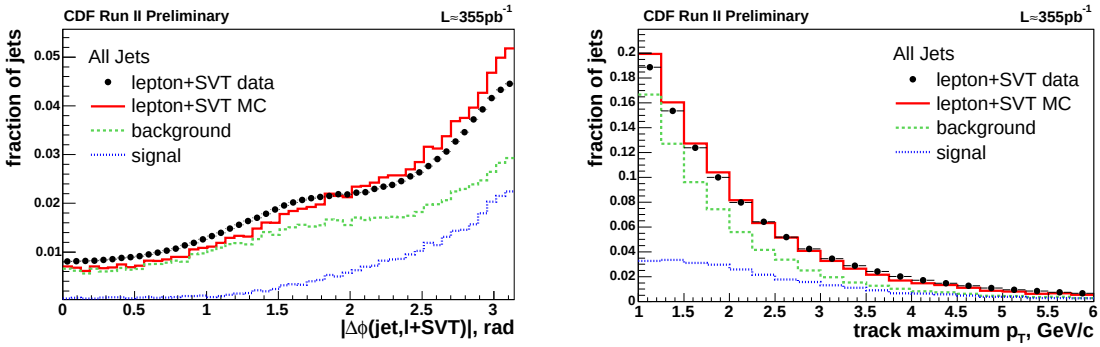

Figure 6.18: Input variables for the bJetNet: angle in the transverse plane between the jet and the B candidate (*left*) and p_T of the track with the maximum transverse momentum in the jet (*right*).

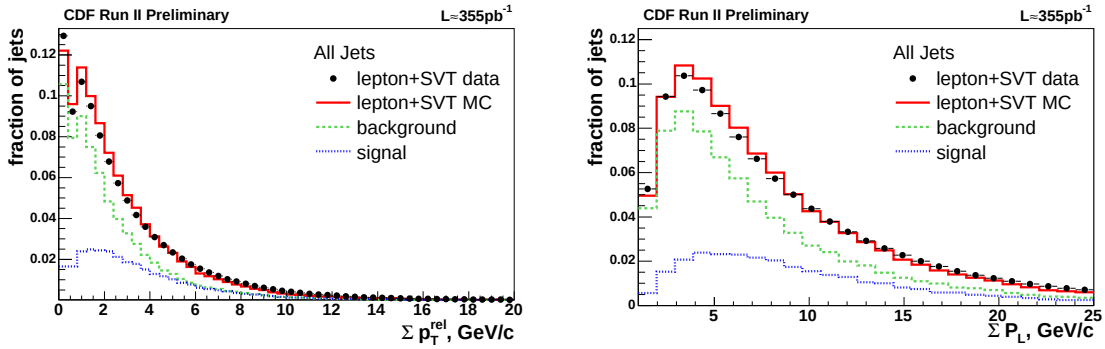

Figure 6.19: Input variables for the bJetNet: $\sum p_T^{rel}$ (*left*) and $\sum p_L$ (*right*).

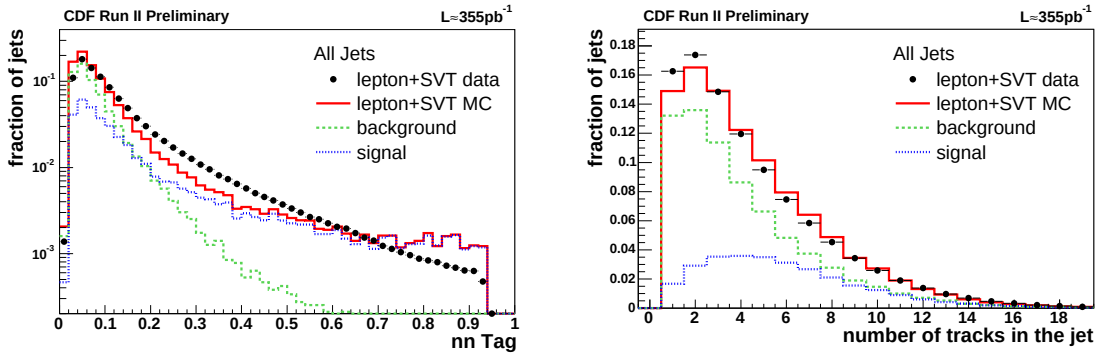

Figure 6.20: Input variables for the bJetNet: T_{nn} (*left*) and number of tracks in the jet (*right*).

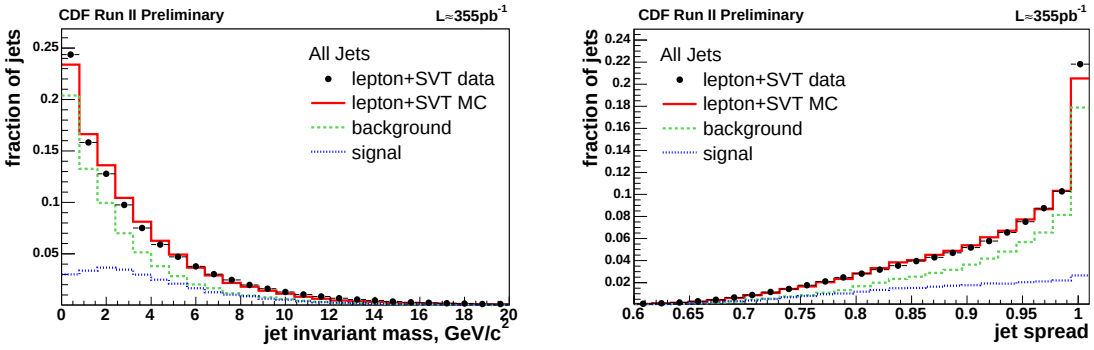

Figure 6.21: Input variables for the bJetNet: jet invariant mass (*left*) and jet spread (*right*).

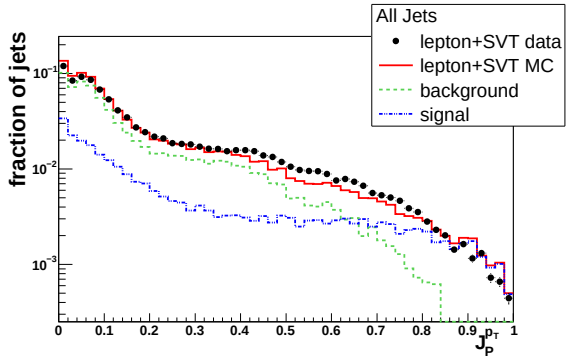

Figure 6.22: Distributions of the jet probability J_P^{pT}. L00 has not been used. The data/Monte Carlo agreement is better than in the case in which L00 hits are added to the tracks (fig. 6.17, left). The computation of the χ^2 as $\sum_i \left((f_i^{data} - f_i^{MC})/\sigma_i^{data}\right)^2$, where the sum is over the bins, f_i is the content of the bin i and σ_i^{data} is the error for the bin i, yielded $\chi^2 = 1.6 \cdot 10^4$ for the distributions of data and Monte Carlo shown in this figure and $\chi^2 = 5.6 \cdot 10^4$ for the distributions in fig. 6.17 (left).

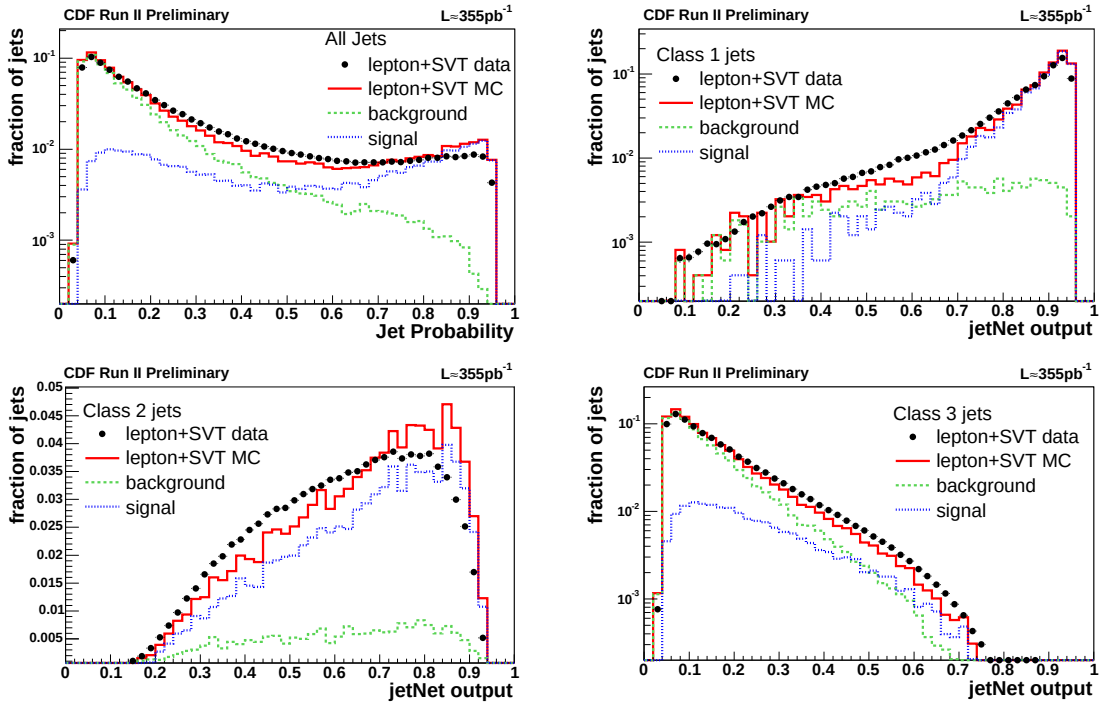

Figure 6.23: Distributions of jet probability for all reconstructed jets (*top left*, for Class 1 jets (*top right*), Class 2 jets (*bottom left*) and Class 3 jets (*bottom right*). The contribution of signal and background jets for Monte Carlo is shown. The data distribution and the inclusive Monte Carlo distribution are normalised to a unit area for each plot.

6.3.2 Performance

The result of the application of the bJetNet to data and Monte Carlo is illustrated in fig. 6.23. Every reconstructed jet has an output value and the bJetNet output distribution for jets in different classes follows the hierarchy of the jet quality, thus suggesting a straightforward jet selection criterion.
The performance of the bJetNet output is compared in fig. 6.24 (left) to that of the IP jet probability introduced in Section 3.4.3. The definitions of efficiency and purity are

$$\text{purity} = \frac{N_{jets}^{signal}(\text{output} > \text{cut})}{N_{jets}^{signal}(\text{output} > \text{cut}) + N_{jets}^{background}(\text{output} > \text{cut})} \tag{6.3}$$

$$\text{efficiency} = \frac{N_{jets}^{signal}(\text{output} > \text{cut})}{N_{jets}^{signal}} \tag{6.4}$$

The bJetNet curve is above the IP jet probability curve, therefore for a given efficiency in selecting signal jets it achieves a higher purity. As a cross check of the separation

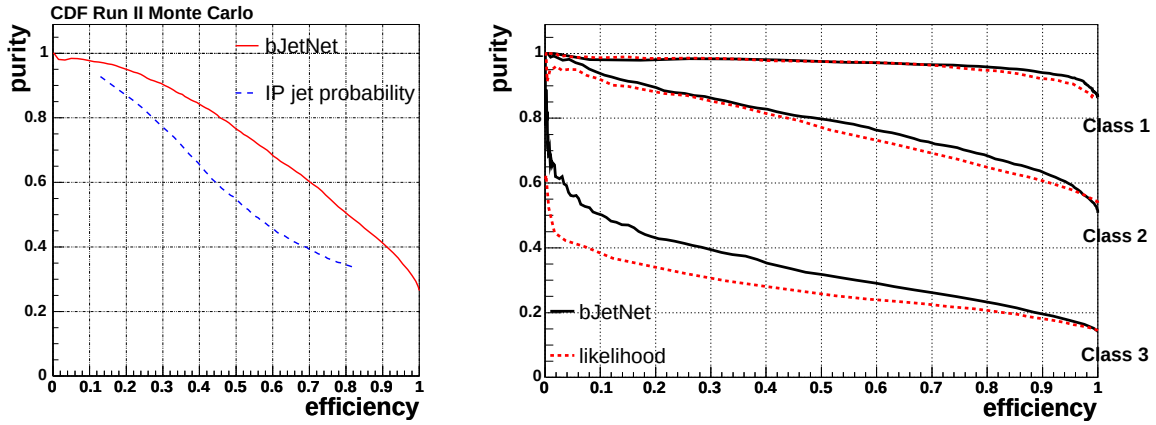

Figure 6.24: *Left:* Performance graph for a cut on the bJetNet output (*solid line*) in simulated events. The comparison with the IP jet probability is shown (*dashed line*). *Right:* bJetNet performance (*solid line*) compared to likelihood (*dashed line*) for the three different classes of jets.

power of the b-jet Neural Network, a b-jet selection likelihood variable was developed using mainly the same input variables of the bJetNet. The correlated variables were excluded and different input sets were defined according to the jet class. The Probability Density Functions have been extracted from the Monte Carlo for signal and background jets. A detailed description of the likelihood method can be found in Appendix E. The jet sample is split into subsamples corresponding to the jet classes and the efficiency and purity of the bJetNet are compared to those of the likelihood variable. Efficiency and purity for class j are defined as

$$\text{purity}^j = \frac{N^j_{signal}(\text{output} > \text{cut})}{N^j_{signal}(\text{output} > \text{cut}) + N^j_{background}(\text{output} > \text{cut})} \tag{6.5}$$

$$\text{efficiency}^j = \frac{N^j_{signal}(\text{output} > \text{cut})}{N^j_{signal}} \tag{6.6}$$

The comparison (fig. 6.24, right) shows that the bJetNet provides better results than the likelihood combination. The improvement comes mostly from the inclusion of correlated variables in the input set.
In the follow up the term *jet probability* will refer to the bJetNet output.

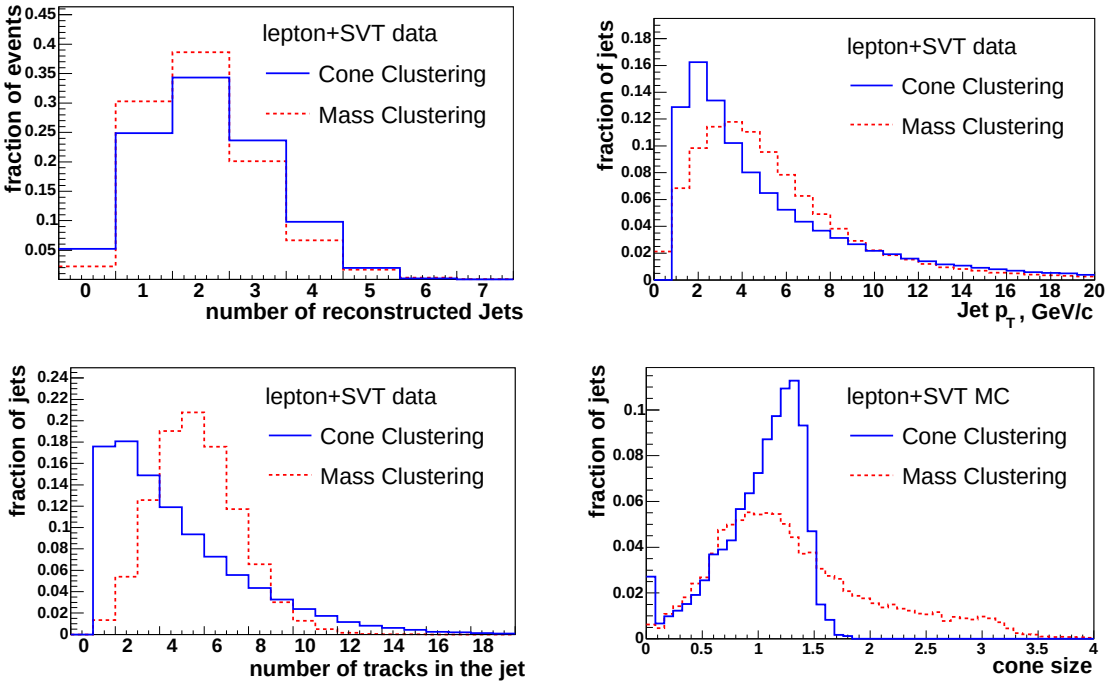

Figure 6.25: Distributions of number of reconstructed jets in the event, jet p_T, number of tracks (data) and cone size for jets (simulation) reconstructed by Cone Clustering (*solid line*) and by Mass Clustering algorithm (*dashed line*).

6.4 *b*-Jet Probability with Mass Clustering

A study [60] has shown that, when the clustering algorithm based on invariant mass (Section 5.6.2) replaces the Cone Clustering in the cut based Jet Charge Tagger, the highest tagging power is about 7% higher than the power achieved with fixed cone jets and is given when the algorithm is run with the parameters listed in table 5.3 and

$$p_T^{seed} = 0.75 \text{ GeV/c}$$
$$x_{min} = 49 \text{ (GeV/c}^2)^2$$

It is interesting then to evaluate the result of the trackNet and the bJetNet with jets reconstructed by the Mass Clustering algorithm. Seed tracks are softer than Cone Clustering seeds ($p_T^{seed} = 1$ GeV/c). Consequently the fraction of events in which at least one jet is reconstructed on the opposite side is larger for Mass Clustering than for Cone Clustering. There tend to be less jets reconstructed by Mass Clustering per event and they contain more tracks than those based on a fixed cone. A comparison between the jets reconstructed by the two algorithms is shown in fig. 6.25. Table 6.3 summarises some jet properties.

	Mass Clustering	Cone Clustering
average number of reconstructed jets	2.2	2.8
average number of tracks in jets	5.2	4.8
average jet p_T	5.4 GeV/c	5.1 GeV/c
fraction of events with 0 jets	2%	5%

Table 6.3: Summary of some properties of jets reconstructed by Cone and by Mass Clustering algorithm.

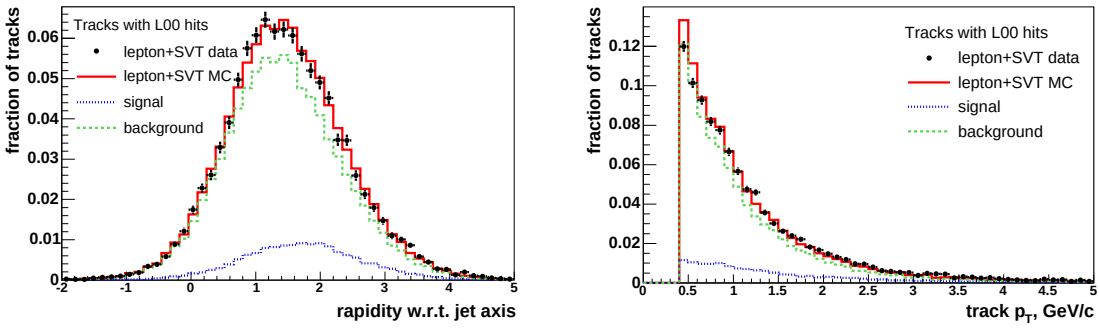

Figure 6.26: Distributions of the rapidity of the track with respect to the jet axis and of the track p_T for tracks with a L00 hit attached to them. Mass Clustering has been used to reconstruct jets.

Since the properties of the jets reconstructed by the two algorithms are different, the variables used for the trackNet and for the bJetNet have different distributions than for the fixed cone case. The training for jets reconstructed with Cone Clustering is then not optimal for jets reconstructed by Mass Clustering. The Neural Networks have been re-trained to suit the new jet distributions. The input variable sets have been left unchanged as well as the setup of the Neural Networks.
The distributions of the input variables which are significantly different with respect to the fixed cone jets are shown in fig. 6.26. Since these variables are not related to the impact parameter information, their distributions for tracks with and without L00 hits are the same, consequently only distributions for tracks with L00 hits are shown. The distribution of the output (fig. 6.27) and the trackNet performance (fig. 6.28) are similar to the results of the trackNet with fixed cone jets (fig. 6.12 and fig. 6.13).
The parameters used for SecVtx in conjunction with Mass Clustering are the same as those of Cone Clustering (table 6.1), except for the cone angle in which SecVtx is allowed to search for tracks to fit in the secondary vertex. This parameter is set to 1000.0 to take into account the broadness of these jets (fig. 6.25, lower right).
The number of jets found in each class is reported in table 6.4. The fraction of jets tagged by SecVtx is the same as for fixed cone jets (see table 6.2). The fraction of

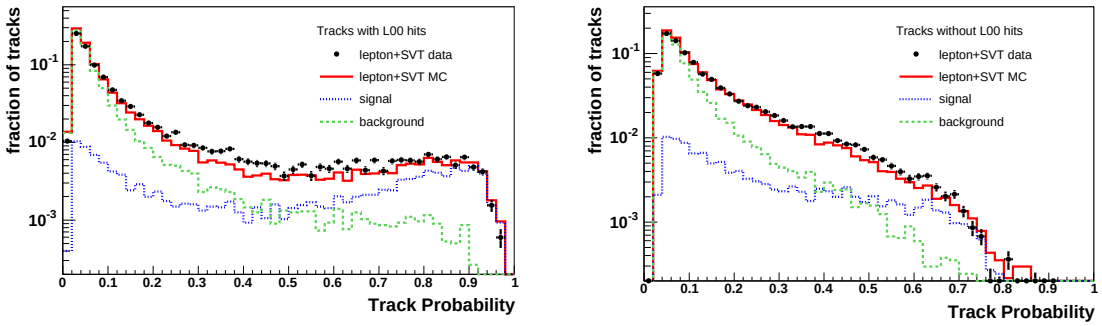

Figure 6.27: Neural Network output for the tracks with L00 hits (*left*) and without (*right*) for the Mass Clustering case.

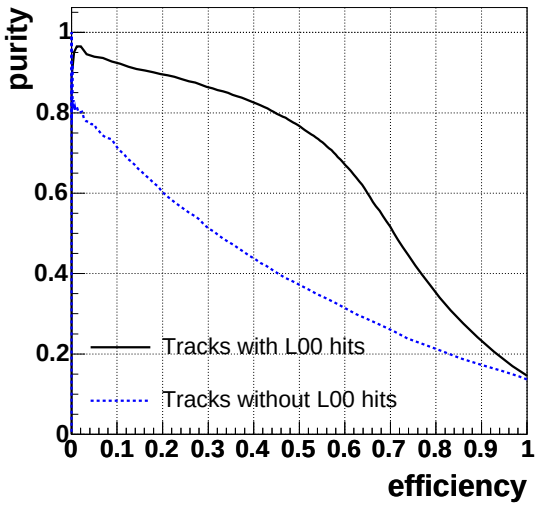

Figure 6.28: Performance graph for the trackNet corresponding to the training with tracks in jets reconstructed by Mass Clustering algorithm.

Class 2 jets is about 2.5% higher. The purity of Class 1 and Class 2 jets is lower.The purity the the whole jet sample is 7.5% higher than the purity of the sample of jets reconstructed by the Cone Clustering algorithm. This is due to the smaller number of jets reconstructed by the Mass Clustering algorithm, since the two algorithms have run on the same Monte Carlo sample.
The distribution of the input variables for the bJetNet which changed with respect to the Cone Clustering case are shown in fig. 6.29. The result of the training and the bJetNet performance graph are shown, respectively, in fig. 6.30 and in fig. 6.31. The distribution of the output is very similar to the distribution in fig. 6.23. The performance graph is compared to that of the bJetNet with Cone Clustering jets to

	Class 1	Class 2	Class 3	Total
e+SVT data				
N jets found	74100	271700	1193000	1538800
fraction	4.8%	17.7%	77.5%	100%
μ+SVT data				
N jets found	96900	366800	1606000	2069700
fraction	4.7%	17.7%	77.6%	100%
Monte Carlo				
N jets found	4270	13320	57650	75240
fraction	5.7%	17.7%	76.6%	100%
signal fraction	86.0%	63.0%	22.5%	36.0%

Table 6.4: Number of jets and fraction of each type of jets found for each type of jets for data and Monte Carlo. The jets are reconstructed with the invariant mass based clustering algorithm. The purity for each subsample in the Monte Carlo is given.

highlight that the new bJetNet performs better only in the efficiency range [0.9,1] and for efficiency lower than 90% it achieves ∼5% lower purity.

In the next chapter the bJetNet trained on jets found by Mass Clustering and the bJetNet trained on jets found by Cone Clustering are incorporated into a Jet Charge Tagger to test the effect of the different jet reconstruction algorithms on b-flavour tagging.

6.5 Summary

A b-track probability variable has been constructed by combining in a Neural Network information sensitive to B decay tracks, e.g. the impact parameter and the rapidity with respect to the jet axis. To take advantage of the higher resolution on the impact parameter for tracks with hits in the innermost layer of the silicon detector, L00, the training of the track probability network has been performed separately on tracks with/without hits in L00. The track probability variable successfully distinguishes B decay tracks from the background in Monte Carlo and in data.
The b-track probability has been employed to define several jet-level variables, which are combined with a Neural Network, resulting in a b-jet probability variable. This has been developed separately for jets with a fixed cone size and for jets with variable cone size. The performance is better on jets with fixed cone size.
The b-jet probability variable is more powerful than a probability based only on the track displacement information and it is going to serve as a b-jet selection criterion in the Jet Charge Tagger.

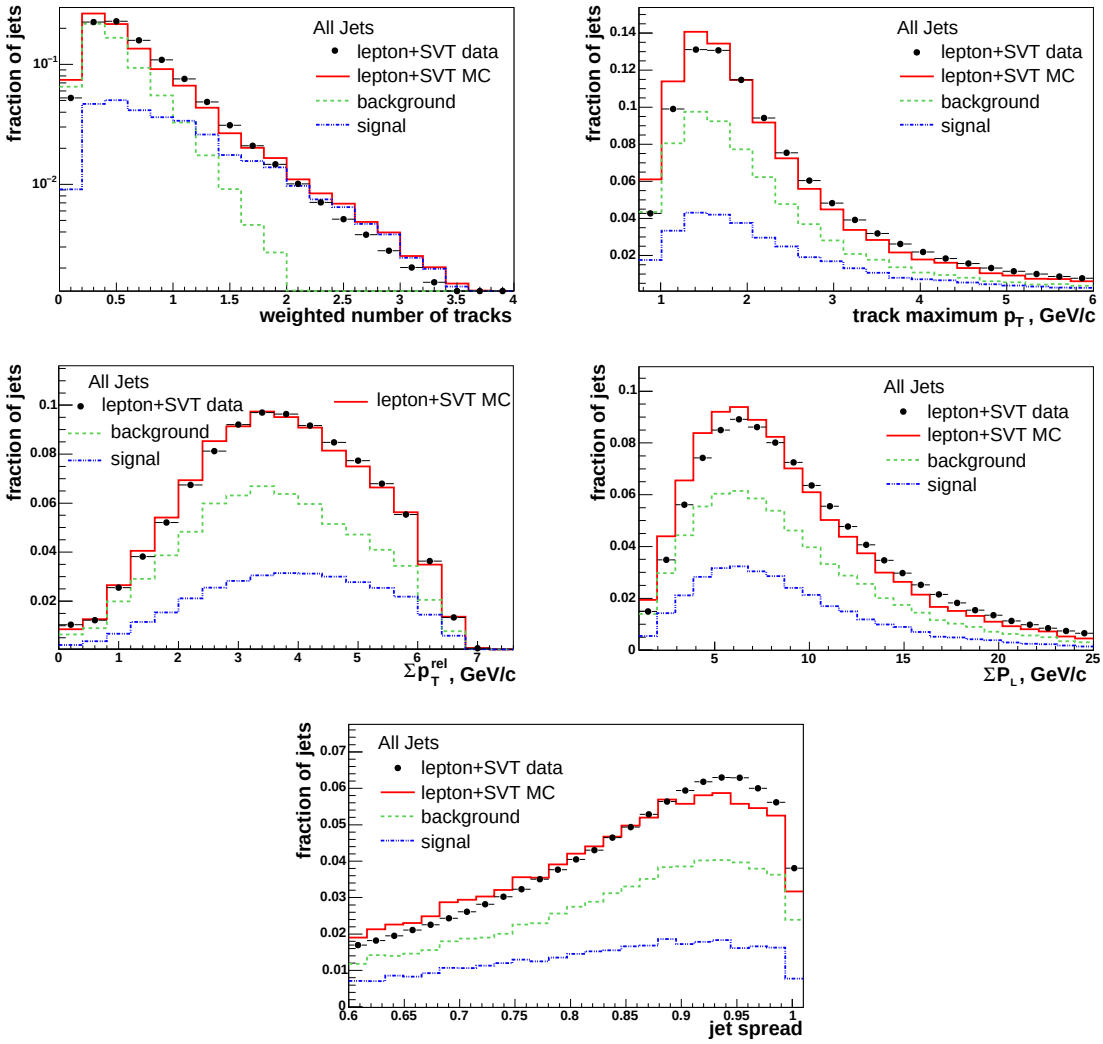

Figure 6.29: Input variables for the bJetNet: weighted number of tracks, p_T of the track with the maximum transverse momentum in the jet, $\sum p_T^{rel}$, $\sum p_L$ and jet spread. The distributions are relative to all jets reconstructed in data and Monte Carlo events with Mass Clustering.

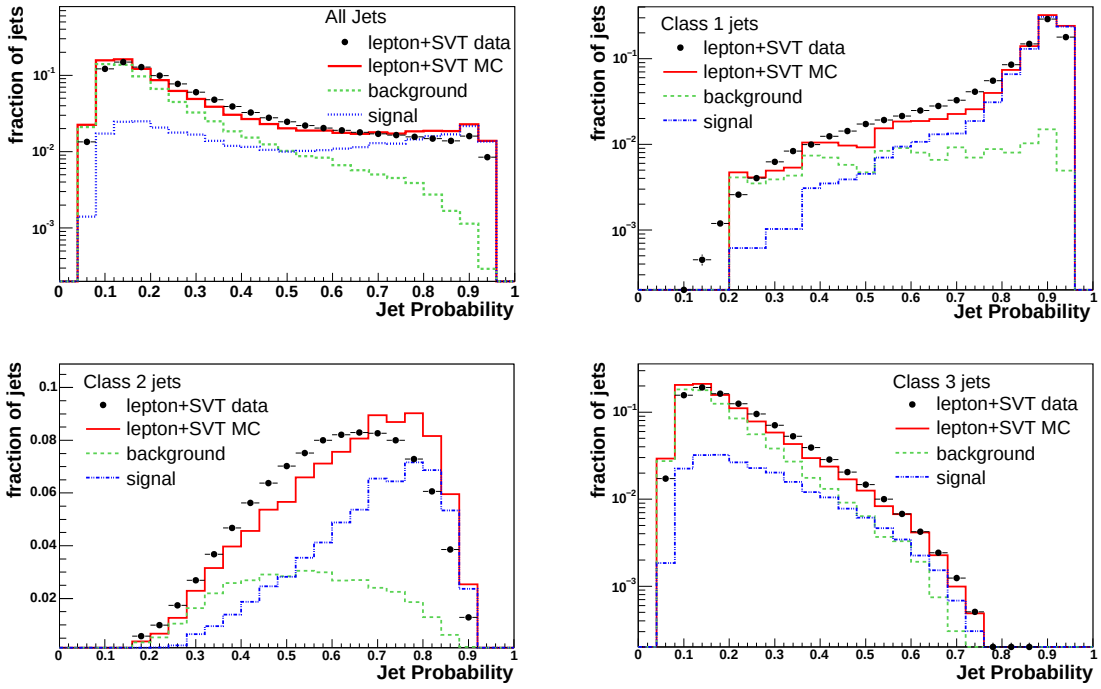

Figure 6.30: Distributions of bJetNet output for all jets (*top left*), for Class 1 (*top right*), Class 2 (*bottom left*) and Class 3 (*bottom right*) jets. The data distribution and the inclusive Monte Carlo distribution are normalised to an area of 1 for each plot. The jets are reconstructed with Mass Clustering.

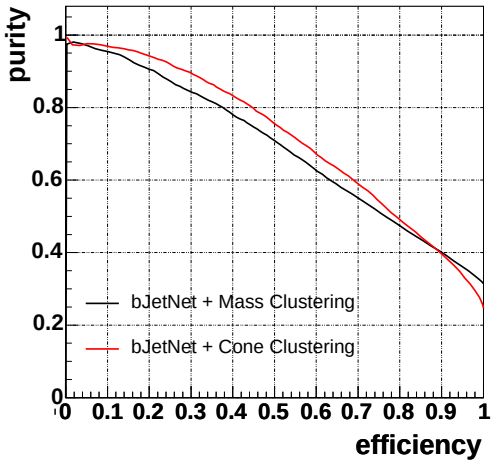

Figure 6.31: Performance graph for a cut on the bJetNet output for jets in simulated events reconstructed using the Mass Clustering algorithm. The curve is compared to the performance of the bJetNet trained with jets reconstructed by the Cone Clustering algorithm.

Chapter 7

Neural Network Based Jet Charge Tagger

The b-jet selection Neural Network variable introduced in the previous chapter is the central tool of the Neural Network based Jet Charge Tagger ("NNJQT"). The correlation of the jet probability to the jet purity is first exploited to select the tagging jet in the event and then to parameterise the tagging power of NNJQT.

7.1 Tagger Setup

The event pre-selection described in Section 6.1 is applied to the data sample; the B candidate is identified as the sum of the momenta of lepton and the SVT track and the tracks on the opposite side are filtered, removing the same side B daughters. The filtered tracks are used to reconstruct jets with the Cone Clustering algorithm.
The flavour on the same side is identified by the charge of the trigger lepton. The Jet Charge Tagger selects a jet on the opposite side with which the flavour of the B candidate can be tagged. In the data sample a jet of Class 2 is preferred to a jet of Class 1 in 1.4% of the events tagged by Class 2 jets. A jet of Class 3 is selected despite of the presence of a Class 2 jet in the event in 2% of the cases and it happens rarely that it has a higher probability than a Class 1 jet (0.2% of the cases).
The tagging jet does not have to satisfy any additional requirement. In particular no cut on the angle in the transverse plane $\Delta\phi$ between the jet and the B candidate is applied. A cut on $\Delta\phi$ has been used in the cut based Jet Charge Tagger [48] for jets with small IP probability and jets with high p_T to remove jets contaminated by the decay products of the same side B. Due to the rejection cut against the same side B decay tracks, the fraction of tagging jets containing of these tracks is very small (see table 7.1), therefore the $\Delta\phi$ cut can be dropped.

Jet type	n tagging jets	n jets with SS B tracks	fraction
Class 1	4335	8	0.18 %
Class 2	10433	19	0.18 %
Class 3	22886	21	0.09%

Table 7.1: Fraction of tagging jets including one or more same side B daughters in the Monte Carlo sample. The contamination is negligible due to the removal of the same side B daughters performed before jet clustering (Section 6.1.2).

7.2 Q_{jet} Computation

The opposite side flavour is given by the sign of the Jet Charge computed for the tagging jet as:

$$Q_{jet} = \frac{\sum_i Q_i \cdot p_{T,i} \cdot (1 + t_i)}{\sum_i p_{T,i} \cdot (1 + t_i)} \tag{7.1}$$

where the index i runs over all the tracks in the jet, Q_i is the track charge, $p_{T,i}$ is the track transverse momentum and t_i is the track probability. The formula 7.1 is equivalent to the one used for the cut based Jet Charge Tagger, with the substitution of the Neural Network probability to the IP probability. The weighting with the track transverse momentum enhances the contribution of the B decay product over the fragmentation tracks.
The distributions of Q_{jet} for positive and negative signed trigger leptons and for the different jet classes are shown in fig. 7.1. The peaks at ± 1 correspond to jets comprising a single track. As expected, the distributions are shifted toward positive (negative) values for negative (positive) signed trigger leptons and the shift is larger for high quality jets.

7.3 Measurement of the Tagging Power $\mathcal{T}$

The NNJQT gives a right sign tag (RS) when Q_{jet} and the trigger lepton charge have the opposite sign, otherwise the tag is wrong (WS). If the tagging jet is not found, the event is not tagged (NT). Since no cuts are applied to the tagging jets, the only events that do not have a tag are those in which the clustering algorithm did not find any jet on the opposite side, i.e. the events in which no track with a transverse momentum larger than 1 GeV/c was found on the opposite side.
The background subtraction with δ_0^{SVT} (Section 5.2.1) is performed on the number of RS, WS and NT events. The procedure to compute the error on the efficiency and the dilution taking into account the background subtraction is described in ref. [63].

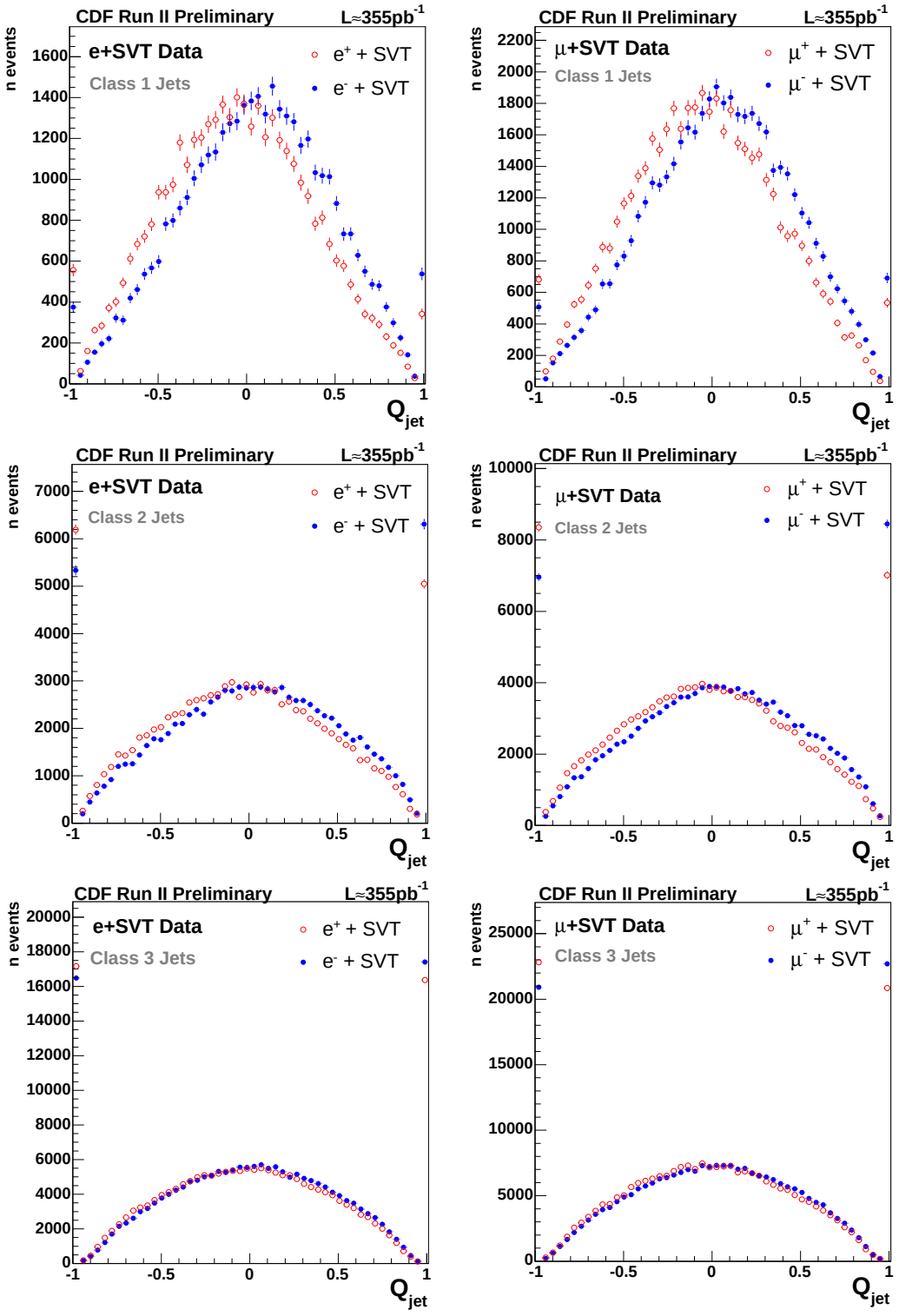

Figure 7.1: Distribution of Q_{jet} for the three jet classes (tagging jets only), for the e+SVT and the μ+SVT sample.

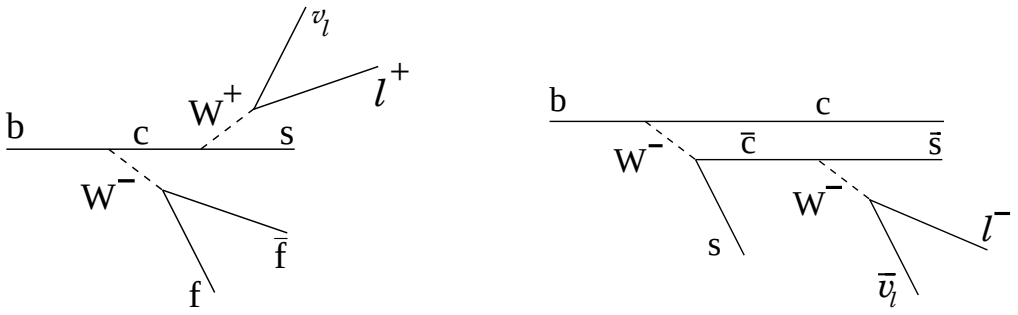

Figure 7.2: Sequential semileptonic decay $b \to c \to \ell$. *Right:* this graph is a particular sequential decay in which a c- and a $\bar{c}$-quark are produced. The lepton originating from the decay of the $\bar{c}$-quark has the right sign correlation with the opposite side flavour.

7.3.1 D_{raw} and D_{true}

The dilution computed according to the formula

$$D = \frac{N_{RS} - N_{WS}}{N_{RS} + N_{WS}} \tag{7.2}$$

is called *Raw Dilution* (D_{raw}). The *True Dilution* (D_{true}) is obtained by scaling D_{raw}

$$D_{true} = \frac{D_{raw}}{s} \tag{7.3}$$

to account for the wrong sign correlation on the same side due to mixing[1] and to sequential semileptonic decays. The scaling factor was computed in ref. [56] based on simulation studies and it is s =0.6412.
In the Monte Carlo sample used for the studies presented in this thesis in 6.1% of the events the trigger lepton is generated in the decay $b \to c \to \ell$ (fig. 7.2). Not all of these leptons have the wrong sign, because of mixing and decays of b-hadrons to double charm (fig. 7.2, right), which could give a lepton of the right sign. The fraction of events with a wrong sign lepton coming from a sequential decay is 5.0%.

7.3.2 Binned Dilution

In the section the Neural Network based Jet Charge Tagger is calibrated on the ℓ+SVT data sample.
The sample of tagging jets is split in subsamples, the true efficiency and the dilution are measured and the partial tagging power is computed for each subsample as the

[1]The fraction of neutral B mesons on the same side is ∼50% and about 17% of them oscillate before decaying [1].

product of the efficiency and the dilution squared. It is reminded that the name "dilution" is misleading, being this a quantity related to the ability of the tagger to give right sign tags. A high dilution value leads to a high tagging power, which is the quantity to maximise. In order to achieve a high tagging power, this is measured on subsamples and summed up to give the combined tagging power $\mathcal{T}$. In fact, if the efficiency is ϵ and the dilution is D on the total sample and if two subsamples can be defined so that one has a higher dilution than the other, simple calculations show that the tagging power measured on the undivided sample $\mathcal{T}_{ave}$ is lower than the sum of the tagging power values measured on each subsample[2]

$$\mathcal{T}_{ave} = \epsilon D^2 \tag{7.5}$$

$$\mathcal{T} = \epsilon_1 D_1^2 + \epsilon_2 D_2^2 \tag{7.6}$$

$$\mathcal{T}_{ave} < \mathcal{T} \tag{7.7}$$

It follows that splitting the calibration sample in subsamples yields a gain in tagging power. Another advantage of this method is that the dilution can be parameterised as a function of the variable used to split the sample, which allows the prediction of the dilution on an event by event basis in a mixing analysis.
For the cut based Jet Charge tagger [47] the sample is split in bins of $|Q_{jet}|$ and jet type. The dilution is expected to be high for a sample of events with charged b-hadrons on the opposite side because they do not mix and the $|Q_{jet}|$ of the corresponding jets is on average larger that the charge of a jet containing a neutral b-meson. Therefore the dilution should depend on $|Q_{jet}|$.
An additional quality information on the tagging jet, i.e. the jet probability variable, is available for the Neural Network based tagger and it is independent from $|Q_{jet}|$ (fig. 7.3). A cut on jet probability rejects events in which the tagging jet is background. Even if a cut might bring a higher dilution, it reduces the efficiency due to the higher number of non-tagged events. A better use of the jet probability variable is to combine it with $|Q_{jet}|$ and split the sample according to the combined variable[3].
The tagging jet sample is divided in 10 bins of $|Q_{jet}| \cdot P_{nn}$ plus a bin for jets containing a single track, for which $|Q_{jet}| = 1$. These jets are kept apart because the flavour information that they give does not come from a weighted average but from a single track charge. The sample splitting is performed separately for different classes of jets,

[2]Given that $\epsilon = \epsilon_1 + \epsilon_2$, it is easy to see that $D = (\epsilon_1 D_1 + \epsilon_2 D_2)/\epsilon$. Taking this relation into account and performing simple algebra, one can find that

$$\epsilon D^2 - \epsilon_1 D_1^2 - \epsilon_2 D_2^2 = -\frac{\epsilon_1 \cdot \epsilon_2}{\epsilon} (D_1 - D_2)^2 \tag{7.4}$$

Since the efficiency and the dilution are positive numbers, the left-hand side of the equation is a negative number if $D_1 \neq D_2$ and zero otherwise. Therefore $\epsilon D^2 < \epsilon_1 D_1^2 + \epsilon_2 D_2^2$ if $D_1 \neq D_2$.

[3]Due to limited statistics, a two-dimensional binning is not possible.

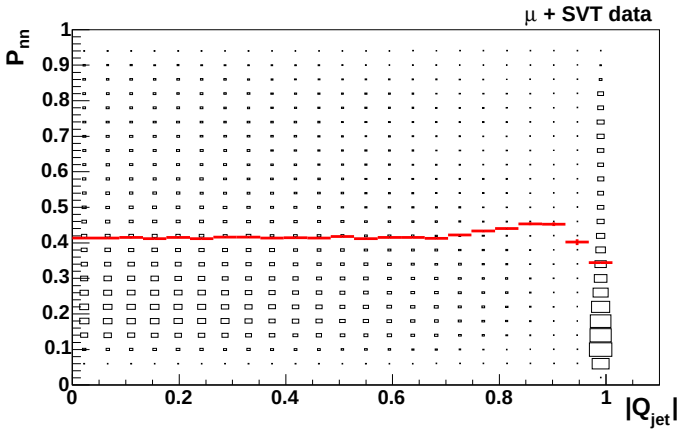

Figure 7.3: Scatter plot of the jet probability P_{nn} and $|Q_{jet}|$. The entries at $|Q_{jet} = 1|$ are due to jets constituted by a single track. The profile of the scatter plot is superimposed.

so in the end there are 33 bins.
The number of RS, WS and NT events and dilution are computed for each bin. Given a jet class c, for each bin k one computes

$$\epsilon_{c,k} = \frac{N^c_{\mathrm{RS},k} + N^c_{\mathrm{WS},k}}{N_{\mathrm{RS}} + N_{\mathrm{WS}} + N_{\mathrm{NT}}} \tag{7.8}$$

where $N^c_{\mathrm{RS},k}$ is the number of right sign tags and $N^c_{\mathrm{WS},k}$ is the number of wrong sign tag for jet type c and bin k. The dilution in the bin, $D_{c,k}$, is measured as

$$D_{c,k} = \frac{N^c_{\mathrm{RS},k} - N^c_{\mathrm{WS},k}}{N^c_{\mathrm{RS},k} + N^c_{\mathrm{WS},k}} \tag{7.9}$$

The dilution is parameterised as a function of $|Q_{jet}|P_{nn}$. The jet purity grows for increasing values of the jet probability P_{nn} (see fig. 6.23), thus affecting the dilution in the same direction. The dilution has a linear dependency on $|Q_{jet}|$, as it has been observed by previous Jet Charge related studies at CDF [48, 64], but the dependency of D on $|Q_{jet}|P_{nn}$ cannot be trivially predicted. The parameterisation of $D(|Q_{jet}|P_{nn})$ is found to be linear (fig. 7.4) and the function $D = a \cdot |Q_{jet}|P_{nn}$ is fitted to the data points. The results of the linear fit for the different jet types, for the e+SVT and the μ+SVT data samples, are listed in table 7.2. As expected, the parameter a depends on the jet type and it reaches the highest values for Class 1 jets, for which a secondary vertex has been identified.

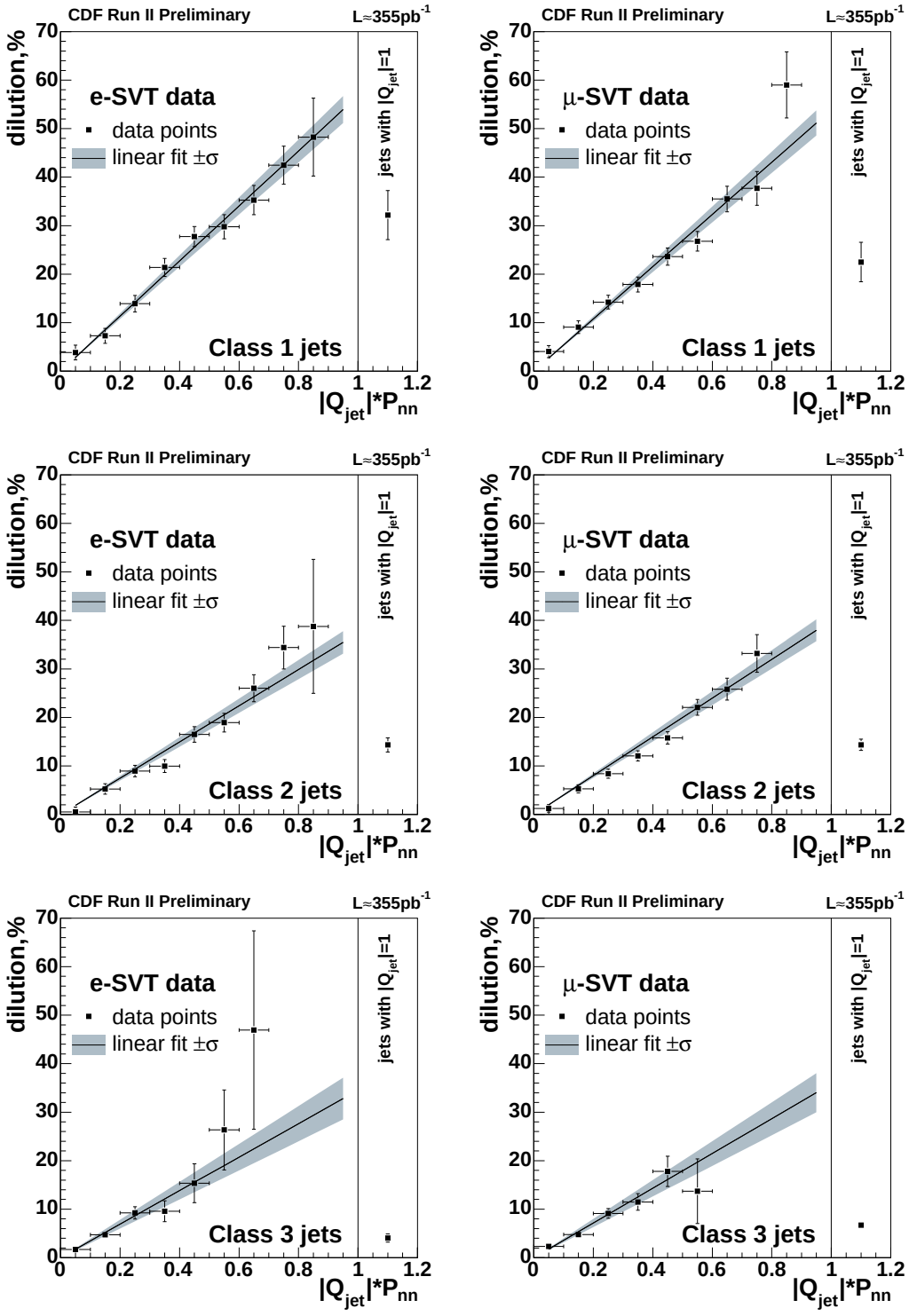

Figure 7.4: Dilution dependency on $|Q_{jet}|P_{nn}$ for tagging jets.

7.3.3 Effective Dilution

In order to quantify the tagger performance with only three figures, instead of a value of $\mathcal{T}_i$ for each bin, the tagging power in the different bins of jet class and $|Q_{jet}|P_{nn}$ is summed up. The total efficiency (ϵ_{tot}), the *effective dilution* (D_{eff}) and the combined tagging power are computed as follows. The *expected dilution* in each bin $\mathcal{D}_{c,k}$ is given by $\mathcal{D} = a \cdot |Q_{jet}|P_{nn}$, except for the last bin ($|Q_{jet}| = 1$), for which the measured value is taken. The error on $\mathcal{D}_{c,k}$ is the error on the expected value, give by the linear fit. The partial tagging power in the bin is $\mathcal{T}_{c,k} = \epsilon_{c,k}(\mathcal{D}_{c,k})^2$. The jet type tagging power is given by the sum over all the bins of the partial tagging power

$$\mathcal{T}_c = \sum_{k=1}^{11} \mathcal{T}_{c,k} \tag{7.10}$$

The jet type efficiency is

$$\epsilon_c = \sum_{k=1}^{11} \epsilon_{c,k} \tag{7.11}$$

and the *effective dilution* for the jet type is

$$D_{c,\text{eff}} = \sqrt{\frac{\mathcal{T}_c}{\epsilon_c}} \tag{7.12}$$

The combined tagging power is

$$\mathcal{T} = \sum_{c=1}^{3} \mathcal{T}_c \tag{7.13}$$

	a	$D(\lvert Q_{jet}\rvert = 1)$, %
e +SVT		
Class 1	0.567 ± 0.029	32.6 ± 5.0
Class 2	0.373 ± 0.024	14.2 ± 1.4
Class 3	0.345 ± 0.045	3.95 ± 0.84
μ +SVT		
Class 1	0.538 ± 0.027	22.5 ± 4.1
Class 2	0.399 ± 0.024	14.3 ± 1.1
Class 3	0.357 ± 0.042	6.71 ± 0.67

Table 7.2: Result of the fit of the function $D = a \cdot |Q_{jet}|P_{nn}$ to data. The measured value of D for single track jets is also given (see fig. 7.4).

Jet type	ϵ %	effective D %	tagging power %
e+SVT data			
Class 1	10.56 ± 0.05	19.75 ± 0.43	0.412 ± 0.018
Class 2	28.20 ± 0.08	11.60 ± 0.33	0.380 ± 0.021
Class 3	56.72 ± 0.11	4.70 ± 0.29	0.125 ± 0.015
combined	95.48 ± 0.15	9.80 ± 0.16	0.917 ± 0.031
μ+SVT data			
Class 1	10.51 ± 0.04	18.35 ± 0.37	0.354 ± 0.014
Class 2	28.76 ± 0.07	12.18 ± 0.29	0.426 ± 0.020
Class 3	56.45 ± 0.09	5.29 ± 0.26	0.158 ± 0.016
combined	95.72 ± 0.12	9.90 ± 0.15	0.938 ± 0.029

Table 7.3: Tagging power of the Neural Network based Jet Charge Tagger. The errors are purely statistical.

and accordingly

$$\epsilon_{tot} = \sum_{c=1}^{3} \epsilon_c \tag{7.14}$$

$$D_{\text{eff}} = \sqrt{\frac{\mathcal{T}}{\epsilon_{tot}}} \tag{7.15}$$

are the total tagging efficiency and the effective dilution.

7.3.4 Results

The tagging power measured on the calibration sample is shown in table 7.3. The measured $\mathcal{T}$ is

$$\begin{array}{ll} (0.917 \pm 0.031)\% & \text{for the } e\text{+SVT data sample} \\ (0.938 \pm 0.029)\% & \text{for the } \mu\text{+SVT data sample} \end{array}$$

The cut based Jet Charge Tagger measures $\mathcal{T} = (0.715 \pm 0.027)\%$ so the overall relative improvement brought by the NNJQT is 29%.
The very high tagging efficiency of NNJQT, about 95%, is due to the absence of cuts on the tagging jet. When the dilution is binned only with $|Q_{jet}|$ the measured $\mathcal{T}$ for NNJQT is 11% lower than the measurement which has just been presented. It is interesting to point out that the tagging power of NNJQT is 12% lower if L00 is not included, because of the lower power of the b-track and b-jet probability variables. The result with dilution binning only in $|Q_{jet}|$ and the result without L00 are listed in table 7.4.

Tagger setup	$\mathcal{T}$ on e+SVT sample	$\mathcal{T}$ on μ+SVT sample
NNJQT + Cone	0.917 ± 0.031 %	0.938 ± 0.029 %
NNJQT + Cone, $D(\|Q_{jet}\|)$	0.848 ± 0.031 %	0.796 ± 0.026 %
NNJQT + Cone, no L00	0.834 ± 0.031 %	0.785 ± 0.027 %

Table 7.4: Comparison of $\mathcal{T}$ of NNJQT+Cone with the case in which the dilution is binned only in $|Q_{jet}|$ and with the case in which L00 is not used. The errors are purely statistical.

As an example of the large gain brought by the splitting of the data sample, the tagging power measured on the e+SVT data sample is only $(0.579 \pm 0.037)\%$ if the sample is split only according to the jet type and $(0.416 \pm 0.033)\%$ if the sample is undivided. Therefore the splitting procedure brings about a factor two of improvement for the tagging power.

7.4 Exclusion of Events Tagged by the Soft Lepton Taggers

The tagging power of NNJQT has been measured on the data sample with the exclusion of the events tagged by the Likelihood based Soft Electron Tagger (SET) or by the Likelihood based Soft Muon Tagger (SMT) [45]. A minimum cut of 0.05 has been applied to the lepton likelihood, similarly to mixing analyses using these taggers. NNJQT can tag every event in which there is at least one track with $p_T > 1$ GeV/c on the opposite side. The lepton selection cuts applied by SET and SMT are stricter, therefore NNJQT tags every event for which SET and SMT have a decision and the overlap of NNJQT with the Soft Lepton Taggers is total.
The measurement of $\mathcal{T}$ on events not tagged by SET or SMT is shown in table 7.5. The efficiency denominator includes all the events in the calibration sample. The dilution fit parameters obtained in this case are charted in table 7.6. The tagging power of NNJQT is lower but is still good and comparable to the power of the best of the two Soft Lepton Taggers (see table 4.1).
The efficiency and the dilution of NNJQT have been evaluated on the overlap sample as well. In this case no binning has been used for NNJQT, nor cuts on Q_{jet} or P_{nn} have been applied due to the limited statistics. The cut on likelihood has been used. The results for the μ+SVT data sample are summarised in table 7.7 and in table 7.8 for e+SVT. These numbers are relevant for the CDF mixing analysis, since the different opposite side taggers are not combined but used in sequence, giving priority to the tagger with the highest dilution.

Jet type	ϵ %	effective D %	tagging power %
e+SVT data			
Class 1	8.75 ± 0.04	17.09 ± 0.44	0.255 ± 0.013
Class 2	25.54 ± 0.08	9.84 ± 0.33	0.247 ± 0.017
Class 3	53.96 ± 0.11	3.76 ± 0.27	0.076 ± 0.011
combined	88.25 ± 0.14	8.10 ± 0.17	0.579 ± 0.024
μ+SVT data			
Class 1	8.67 ± 0.04	15.16 ± 0.37	0.199 ± 0.010
Class 2	25.99 ± 0.06	10.19 ± 0.29	0.270 ± 0.015
Class 3	53.61 ± 0.09	4.16 ± 0.23	0.093 ± 0.010
combined	88.28 ± 0.11	7.98 ± 0.15	0.562 ± 0.021

Table 7.5: Summary of efficiency, dilution and tagging power measured on data for the Jet Charge Tagger based on Neural Network jet selection with the exclusion of events tagged by the Soft Lepton Tagger. The errors are purely statistical.

	a	$D(\|Q_{jet}\|=1)$, %
e +SVT		
Class 1	0.498 ± 0.030	20.6 ± 4.4
Class 2	0.314 ± 0.024	13.3 ± 1.2
Class 3	0.280 ± 0.043	5.83 ± 0.68
μ +SVT		
Class 1	0.448 ± 0.027	26.6 ± 5.5
Class 2	0.330 ± 0.022	12.9 ± 1.5
Class 3	0.271 ± 0.034	3.10 ± 0.85

Table 7.6: Result of the fit of the function $D = a \cdot |Q_{jet}|P_{nn}$ to data when events tagged by the Soft Lepton Taggers are excluded.

Subsample	ϵ %	D %	tagging power %
NNJQT with SMT			
overlap	4.41 ± 0.03	24.11 ± 0.93	0.256± 0.020
decisions agree	3.17 ± 0.02	37.2 ± 1.1	0.440± 0.025
decisions disagree	1.24 ± 0.01	9.7 ± 1.8	0.012± 0.004
NNJQT with SET			
overlap	3.23 ± 0.02	19.2 ± 1.1	0.119± 0.014
decisions agree	2.23 ± 0.02	28.0 ± 1.3	0.175± 0.016
decisions disagree	1.00 ± 0.01	0.5 ± 2.1	–
NNJQT with SMT or SET			
overlap	7.45 ± 0.03	22.20 ± 0.72	0.367± 0.024
decisions agree	5.47 ± 0.03	32.81 ± 0.83	0.589± 0.030
decisions disagree	1.97 ± 0.02	7.2 ± 1.4	0.010± 0.004

Table 7.7: Measurement of ϵ and D of NNJQT on the μ+SVT data sample only for events that were tagged also by the Soft Lepton Tagger. For the computation of D no binning has been used. The efficiency is normalised to the total number of events in the sample, as in table 7.3.

Subsample	ϵ %	D %	tagging power %
NNJQT with SMT			
overlap	4.21 ± 0.03	23.5 ± 1.1	0.233 ± 0.016
decisions agree	3.02 ± 0.03	37.8 ± 1.3	0.431 ± 0.030
decisions disagree	1.19 ± 0.02	12.6 ± 2.2	0.019 ± 0.007
NNJQT with SET			
overlap	3.19 ± 0.03	19.4 ± 1.38	0.119 ± 0.017
decisions agree	2.25 ± 0.02	28.2 ± 1.6	0.179 ± 0.020
decisions disagree	0.93 ± 0.02	2.0 ± 2.7	–
NNJQT with SMT or SET			
overlap	7.23 ± 0.04	21.75 ± 0.89	0.342 ± 0.028
decisions agree	5.34 ± 0.03	32.6 ± 1.0	0.569 ± 0.035
decisions disagree	1.89 ± 0.02	8.9 ± 1.8	0.015 ± 0.006

Table 7.8: Measurement of ϵ and D of NNJQT on the e+SVT data sample only for events that were tagged also by the Soft Lepton Tagger. For the computation of D no binning has been used. The efficiency is normalised to the total number of events in the sample, as in table 7.3.

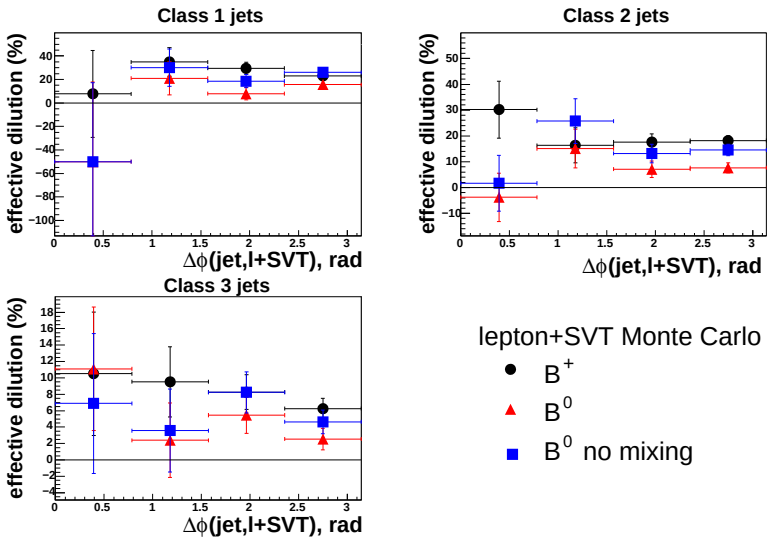

Figure 7.5: Effective dilution in bins of the angle in the transverse plane between the tagging jet and the B candidate for NNJQT with Cone Clustering in simulated events. Monte Carlo truth is used to identify the flavour of the same side B.

7.5 Further Studies

7.5.1 $D(B^0)$ and $D(B^+)$

One of the major improvements with respect to the cut based Jet Charge Tagger comes from the elimination of the cut on the $\Delta\phi$ angle between the tagging jet and the B candidate. By dropping this cut it is possible that the leading fragmentation track on the same side is included more often in the tagging jet on the opposite side. The study in Section 5.5.1 indicates that the tracks originating from the fragmentation of a B cannot be enclosed in a cone and excluded in a simple manner. The charge correlation of the leading fragmentation track to the same side B flavour is opposite for B^0 and B^+, so if the tagging jets include these tracks, the measured dilution of the Jet Charge Tagger should be different depending on whether the same side B is a charged or a neutral meson.
The flavour of the same side B has been identified with the Monte Carlo truth, therefore no correction factor for the raw dilution has been applied. The dilution is measured in bins of $\Delta\phi$ separately for B^+, for all B^0 and for unmixed B^0. The tagger setup is NNJQT with Cone Clustering. The result is shown in fig. 7.5. The dilution is slightly lower for B^0 because of mixing on the same side, for which no correction is applied. It is interesting to compare B^+ and unmixed B^0 distributions, since, if the same side fragmentation tracks do not affect the opposite side, the measured dilution should be the same in the two cases. For Class 1 and Class 2 jets the dilution of

unmixed B^0 is lower in the first $\Delta\phi$ bin than in the other bins. The errors are rather large and make it a difficult task to understand the size of the bias, if there is any significant one. A test with higher statistics would be needed for further understanding. In addition, the fragmentation model in the Monte Carlo is not tuned, therefore the study should be repeated with events simulated with the fragmentation function that fits the data best.
It has to be noted that the dilution measured on simulated events is larger than the dilution measured on data. The simulation gives a good description of the opposite side of the event, but the flavour information does not completely match to the flavour observed in data. Further investigation is needed to explain this difference.

7.5.2 Alternative Jet Charge Tagger Setup

The results shown so far have been obtained using the track probability optimised separately on tracks with/without hits in L00. The tagging power has also been evaluated in conjunction with the trackNet trained on three subsamples (tracks with L00 hits, tracks with hits in the silicon detector but not in L00, tracks with COT hits only). The measurement yielded a tagging power $\sim$ 1.5% larger, which is not significant. The results are given in Appendix G.2.
The possibility of selecting the tagging jet with the likelihood variable (Appendix E) has been considered and a Jet Charge Tagger based on the likelihood variable has been developed (Appendix F). The measured tagging power of the likelihood based Jet Charge Tagger has been found to be on average on the data samples $\sim$10% lower than the power of the Neural Network based Jet Charge Tagger. The reason is that the b-jet likelihood variable is not as powerful as the bJetNet in separating signal and background jets.
An additional setup considered for the Jet Charge Tagger included the use of Mass Clustering. Also in this case the measured tagging power is lower, about $\sim$11% on average. The setup and the results are discussed in the next section.

7.5.3 Jet Charge Tagger with Mass Clustering

The performance of the Jet Charge Tagger has been evaluated when the opposite side jets are reconstructed with Mass Clustering with the optimised values for p_T^{seed} and x_{min} (see Section 6.4).
The Jet Charge Tagger has been run on events in which jets have been reconstructed on the opposite side by Mass Clustering algorithm. For this purpose the trackNet and bJetNet trained with these jets have been employed. The tagger performance has been evaluated with the method described as in Section 7.3.
The fraction of jets selected with the jet probability variable and contaminated by

Jet type	n tagging jets	n jets with SS B tracks	fraction
Class 1	4139	2	0.04%
Class 2	11211	25	0.22%
Class 3	24126	37	0.15%

Table 7.9: Fraction of tagging jets including one or more same side B daughters falling on the opposite side in the Monte Carlo sample.

Jet type	ϵ,%	effective D, %	tagging power %
e+SVT data			
Class 1	9.92± 0.05	20.24± 0.51	0.406± 0.021
Class 2	29.80± 0.09	10.69± 0.33	0.341± 0.021
Class 3	58.36± 0.11	4.37± 0.28	0.112± 0.015
combined	98.08± 0.15	9.36± 0.18	0.859± 0.033
μ+SVT data			
Class 1	9.87± 0.04	18.24± 0.44	0.328± 0.016
Class 2	30.27± 0.07	10.48± 0.28	0.332± 0.018
Class 3	58.09± 0.09	4.41± 0.23	0.113± 0.012
combined	98.23± 0.12	8.86± 0.15	0.774± 0.027

Table 7.10: Summary of efficiency, dilution and tagging power measured on data for the Jet Charge Tagger based on Mass Clustering and Neural Network jet selection.

same side B decay products (table 7.9). There is no significant difference from the numbers obtained for Cone Clustering (table 7.1).
Table 7.10 shows the tagging power of NNJQT with Mass Clustering. The measured $\mathcal{T}$ is 6% to 17% lower than NNJQT with Cone Clustering (table 7.3) and the loss is mainly due to a ∼1% lower effective dilution for Class 2 and Class 3 jets. Class 1 jets have a lower efficiency which is compensated by a slightly higher dilution. The combined tagging efficiency is ∼2.5% higher than NNJQT with Cone Clustering. The improvement is explained by the higher efficiency in jet reconstruction of Mass Clustering due to the lower p_T cut on seed tracks. The jet type with the highest relative efficiency improvements is Class 2.

The use of jets produced by the Mass Clustering algorithm does not bring an improvement in the tagging power. In Run I a Jet Charge Tagger based on Mass Clustering proved to be more powerful than a Tagger based on Cone Clustering when applied to events with low p_T of the B-candidate [59]. In order to understand if the superiority of Cone Clustering in this case is caused by the higher p_T of the B candidate, a test

is performed on the Monte Carlo sample. The simulated events are split in a low and in a high bin of true B candidate transverse momentum[4]. The reason for the usage of simulated events and of the true momentum of the B-candidate is that in the analysis in Run I the B candidates were reconstructed in exclusive channels. The tagging power is measured in the two p_T-bins for NNJQT+Cone and NNJQT+Mass. The flavour of the B candidate is taken from the Monte Carlo truth to determine the true dilution. The Monte Carlo predicts that the performance of NNJQT+Mass is 0.7% worse in the high p_T-bin than in the low p_T-bin, so there is no significant variation. NNJQT+Cone has a tagging power 15% larger in the high p_T-bin than in the low p_T-bin. The significance for the effect is only $\sim 1\sigma$.

7.6 Summary

The Neural Network based Jet Charge Tagger has been presented in this chapter. It uses the b-jet probability P_{nn} to select the tagging jet on the opposite side and the b-track probability as a weight for the computation of the jet charge Q_{jet}. The tagging power has been measured on the ℓ+SVT data sample, which has been split is several subsamples corresponding to bins of the quantity $|Q_{jet}|P_{nn}$ and jet types. The partial tagging power has been measured in each subsample and the sum over the subsamples gave the combined tagging power $\mathcal{T}$, which is larger than the tagging power measured on the undivided sample. The final result is

$$\begin{aligned} \mathcal{T} &= (0.917 \pm 0.031)\% \quad \text{for the } e\text{+SVT data sample} \\ \mathcal{T} &= (0.938 \pm 0.029)\% \quad \text{for the } \mu\text{+SVT data sample} \end{aligned}$$

The tagging power measured with alternative setups, e.g. with a likelihood based b-jet selection or with jets reconstructed by the Mass Clustering algorithm, has been proven to be worse.
The performance of the Neural Network based Jet Charge Tagger was found to be about 30% better than the performance of the cut based Jet Charge Tagger.
A parameterisation has been provided for the evaluation of the tagger dilution on an event-by-event basis in B_s^0 oscillation analyses.

[4]Events with B candidate $p_T < 14$ GeV/c and events with B candidate $p_T > 14$ GeV/c. The cut is chosen to ensure the same statistics in the two bins.

Chapter 8

Conclusion and Outlook

A Jet Charge Tagger algorithm for b-flavour tagging for the measurement of Δm_s at CDF has been presented. The tagger is based on a b-track probability variable and a b-jet probability variable, both obtained by combining the information available in $b\bar{b}$ events with a Neural Network.
The tagging power measured on data is

$$\begin{array}{ll} 0.917 \pm 0.031\% & e\text{+SVT sample} \\ 0.938 \pm 0.029\% & \mu\text{+SVT sample} \end{array}$$

which is $\sim$30% larger than the cut based Jet Charge Tagger employed for the B_s^0 mixing analysis presented by CDF at the Winter Conferences 2005. The improved power of the tagger is due to the selection of the b-jet with a Neural Network variable, which uses correlated jet variables in an optimal way.
The development of the track and jet probability has profited from studies performed on simulated events, which allowed to understand better the features of $b\bar{b}$ events.
For the first time in the CDF B group a Monte Carlo sample comprising flavour creation and additional $b\bar{b}$ production processes has been examined and compared to Run II data. It has been demonstrated that a Monte Carlo sample with only flavour creation $b\bar{b}$ production processes is not sufficient to describe $b\bar{b}$ data collected at CDF. The sample with additional processes introduced in this thesis is thus essential for tagging studies. Although the event description is satisfactory, the flavour information in the Monte Carlo sample differs with respect to data. This difference needs to be clarified by further studies.
In addition, the track and the jet probabilities are the first official tools based on Neural Networks for B-Physics at CDF. They have proven that the simulation is understood to such an advanced level that Neural Networks can be employed. Further work is going on in this direction: a Soft Electron and a Soft Muon Tagger based on Neural Networks are under development as of now.

Several possible tagger setups have been studied and the Jet Charge Tagger reached a high level of optimisation. A further improvement of the tagging power can be achieved by combining the opposite side taggers in a single one, i.e. including particle identification in the track probability. A change of perspective might bring tagging at CDF to a higher performance: the traditional jet clustering could be abandoned in favour of a track-based tag on the opposite side. This approach was successfully pursued in the DELPHI experiment with the BSAURUS project [5]. A similar strategy is currently under investigation for the CDF experiment. The first studies on simulation are encouraging [65].
The presented Jet Charge Tagger marks the advent of new flavour tagging techniques at CDF and it is going to greatly enhance the ongoing Δm_s analysis.

Appendix A

True Asymmetry and Measured Asymmetry

Consider the measurement of an asymmetry which uses a tagger with efficiency ϵ and dilution D to make a binary decision: the event is either of type a or of type b. The measured asymmetry is defined as

$$A_{meas} = \frac{N_a - N_b}{N_a + N_b} \tag{A.1}$$

where N_a (N_b) is the number of events tagged as type a (b). The true asymmetry in the data is given by

$$A_{true} = \frac{N_a^0 - N_b^0}{N_a^0 + N_b^0} \tag{A.2}$$

where N_a^0 (N_b^0) is the number of true events of type a (b) in the sample. The efficiency of the flavour tag is given by

$$\epsilon = \frac{N_a + N_b}{N_a^0 + N_b^0} \tag{A.3}$$

which is the numbers of tagged events over the number of events before the tagging. The number of tagged events of each type is

$$N_a' = \epsilon N_a^0 \tag{A.4}$$

$$N_b' = \epsilon N_b^0 \tag{A.5}$$

The events tagged as type a are either correctly tagged true a-type events or wrongly tagged b-type events, and similarly for the events tagged as type b. Consequently one can use the probability for the tagger to give a right sign, P_{RS} (eqn. 4.14), or a wrong sign tag, P_{WS} (eqn. 4.15), to write:

$$N_a = N_a' P_{RS} + N_b' P_{WS} \tag{A.6}$$

$$N_b = N'_b P_{RS} + N'_a P_{WS} \tag{A.7}$$

and substituting in eqn. A.1 one obtains

$$A_{meas} = \frac{N'_a P_{RS} + N'_b P_{WS} - N'_b P_{RS} - N'_a P_{WS}}{N_a + N_b} \tag{A.8}$$

and using $N_a + N_b = \epsilon(N_a^0 + N_b^0)$ from eqn.

$$A_{meas} = \frac{(N'_a - N'_b)(P_{RS} - P_{WS})}{\epsilon(N_a^0 + N_b^0)} = \frac{\epsilon(N_a^0 - N_b^0)(P_{RS} - P_{WS})}{\epsilon(N_a^0 + N_b^0)} \tag{A.9}$$

The factor ϵ cancels out and from eqn. 4.14 and 4.15 it follows

$$P_{RS} - P_{WS} = D \tag{A.10}$$

$$A_{meas} = D\frac{N_a^0 - N_b^0}{N_a^0 + N_b^0} = D \cdot A_{true} \tag{A.11}$$

which leads to eqn. 4.16.
The form 4.17 for the statistical uncertainty on A_{true} is computed by propagating the errors $\sigma^2_{N_a} = N_a$ and $\sigma^2_{N_b} = N_b$, considering that N_a and N_b are not independent numbers since $N = N_a^0 + N_b^0$, thus N is fixed, and the relation $N_a + N_b = \epsilon \cdot N$ holds.

Appendix B

Production of the Monte Carlo Samples

B.1 Flavour Creation Sample

The Monte Carlo sample with flavour creation only has been generated, simulated and reconstructed at the GridKa computing facility [66].
The sample has been generated with Pythia v6.2 with the option `msel set 5`, that is $b\bar{b}$ flavour creation, with non-zero b-quark mass in the leading order matrix elements. The Bowler fragmentation model has been used, which is the default for Pythia. The decay package used is QQ [67].
The version of CDF software used to produce this sample is 5.3.0. The `cdfSim` executable was a custom version based on 5.3.0 and including the latest version of ToF reconstruction available at the time of the generation (end of April 2004). The set of steering parameters is the default available for 5.3.0 release. The flag `cdfSim_SI_Matching` has been set to 1 in the `cdfSim` steering file in order to keep all the data banks needed for track matching to Monte Carlo particles.

B.2 All $b\bar{b}$ Production Processes Sample

The Monte Carlo sample containing all the $b\bar{b}$ production processes has been produced on the Fermilab CAF by using the `nbot90` dataset [68]. The latter consists of 25M events containing only the four vector quantities of the generated particles. The events have been generated with Pythia v6.2 with the option `msel set 1`, i.e. the production of QCD processes at high transverse momentum. The fragmentation function used is the default, i.e. the Bowler fragmentation function. A filter has been applied to select only events containing at least a b- or a $\bar{b}$-quark with transverse momentum

$p_T > 4$ GeV/c and pseudorapidity $|\eta| < 1.5$. The four vector events have been processed by EvtGen [69] and simulated.
The ℓ+SVT trigger has been simulated and the events surviving after the filter have been reconstructed. Since the trigger filter drastically reduces the initial number of events, the `nbot90` sample has been processed by the decay package and simulated for several iterations. The probability that the same initial event fires the trigger twice is very small. The duplicated events, which amount to a 1% of the total number of simulated events, have been tabulated so that it is possible to exclude them in the DHInput module.
The version of CDF software used to produce this sample is 5.3.3. The `cdfSim` executable has been compiled in order to include the latest version of ToF reconstruction available (`toftag2`). The default parameters of the 5.3.3 release have been used for simulation and reconstruction. Also for this sample the flag `cdfSim_SI_Matching` has been set to 1 in the `cdfSim` steering file.

Appendix C

Optimisation of the Track Probability

The NeuroBayes® package [70] has been used for every Neural Network described in this thesis. The setup of the Neural Network consists of three layers (see fig. C.1):

- input layer: each node in this layer corresponds to an input variable.
- hidden layer: contains the intermediate nodes that correspond to linear combinations of the input nodes
- output layer: the output nodes correspond to the answer given by the Network. In case the decision is of the type yes/no there is only one node in the output layer

In general the aim is to teach the Network to learn the correlation between the input variables and a variable called *target* (e.g. the Monte Carlo truth), and to find the combination of input variables that has the highest correlation to the target.
The *training* is the procedure that assigns a weight to each connection between an intermediate node and an input node and each connection between an intermediate node and an output node and finds the set of weights that gives the smallest difference between the Network output and the target. When the optimal set of weight has been found, the training is complete. The set of weights, called *expertise*, is later used to extract the output given a set of input variables.
The Neural Network output for a binary decision is a float distributed between -1 and $+1$, where -1 means that answer is "no" without any doubt and 1 means that the answer is certainly "yes". When the Network is properly trained the purity of the signal is proportional to the Neural Network output scaled to the interval $(0., 1.)$. This allows one to interpret the Neural Network output as a probability.

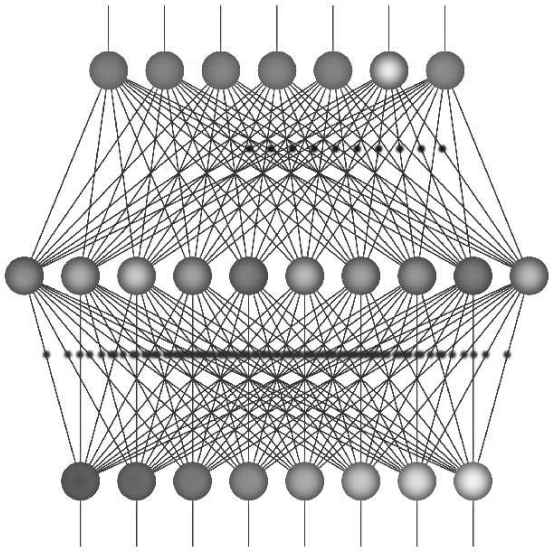

Figure C.1: Scheme of the three layers of a Neural Network and the connections among the nodes in the different layers [71].

target	d_0^{signed}/σ	track p_T	$\lvert d_0 \rvert$	PV	ΔR	p_T^{jet}	y	
	3	2	1	4	5	6	7	sorting
100.0	22.7	15.1	40.7	-20.3	5.2	1.1	10.6	target
	100.0	3.6	14.7	-7.1	2.2	3.3	2.0	d_0^{signed}/σ
		100.0	-9.4	22.6	-0.3	16.0	49.9	track p_T
			100.0	-25.5	2.9	-0.8	-2.1	$\lvert d_0 \rvert$
				100.0	-0.2	-7.5	9.3	PV
					100.0	25.8	7.3	ΔR
						100.0	-15.8	p_T^{jet}
							100.0	y

Table C.1: Correlation matrix of input variables after preprocessing for tracks with a L00 hit. The numbers given are in %.

C.1 Network Training

The trackNet setup has 8 nodes in the input layer (7 input variables plus a bias node, which is needed for technical reasons), 6 nodes in the intermediate layer and one output node.
The training sample consists of 37k signal tracks and 203k background tracks for the tracks with L00 hits and 21k signal tracks and 126k background tracks for the tracks

target	d_0^{signed}/σ	track p_T	$\lvert d_0\rvert$	PV	ΔR	p_T^{jet}	y	
	3	2	1	4	5	6	7	sorting
100.0	14.6	14.9	17.0	-11.7	4.4	1.4	10.2	target
	100.0	3.1	6.4	-4.3	1.6	3.0	1.7	d_0^{signed}/σ
		100.0	-18.0	24.9	2.0	18.6	50.4	track p_T
			100.0	-28.8	1.3	-3.3	-9.0	$\lvert d_0\rvert$
				100.0	0.3	-7.0	9.8	PV
					100.0	26.5	9.7	ΔR
						100.0	-11.4	p_T^{jet}
							100.0	y

Table C.2: Correlation matrix of input variables after preprocessing for tracks without L00 hits. The numbers given are in %.

without L00 hits, equivalent to all of the ℓ+SVT Monte Carlo sample.
The signal to background ratio used for the training is the natural mix found in simulated events. Every track in a jet is used to train the trackNet.
NeuroBayes® processes the input variables, computes the correlation matrix and sorts the variables in order of importance. The importance of a variable is proportional to the information loss when the variable is removed from the set of input nodes.
The correlation matrices for the sample of tracks with L00 hits is shown in table C.1 and for tracks without L00 hits in table C.2. The numbers in the row after the name of the variables indicate the importance of the variable in the corresponding column. The following row shows the correlation of each input variable to the signal. The negative numbers indicate anti-correlation of the variable to the signal. The correlation of each variable to the signal is relatively small, being the maximum 40% for $|d_0|$ in the case of tracks with L00 hits. It would not be possible to obtain a powerful separation of signal and background by using only one of the variables.
As an example on how to read the correlation matrices, consider the correlation between the primary vertex flag, PV, and the impact parameter, $|d_0|$. The PV flag is 1 if the track has been used by VxPrim to fit the primary vertex and it is 0 if the track has not been used. If the primary vertex in the event has not been found then the flag is equal to 2. VxPrim applies a very loose cut on $|d_0|$ to select tracks and fits the primary vertex with an iterative procedure that removes the tracks that gives the highest contribution to the fit χ^2. The tracks that are more likely to be excluded are those with a large displacement, so at the end of the fit procedure the tracks in the primary vertex, which is flagged with 1, have an average $|d_0|$ smaller than the tracks excluded from the primary vertex fit, flagged with 0. In this sense the PV and the $|d_0|$ variables are anti-correlated.
The track impact parameter is the most sensitive variable to the track coming from

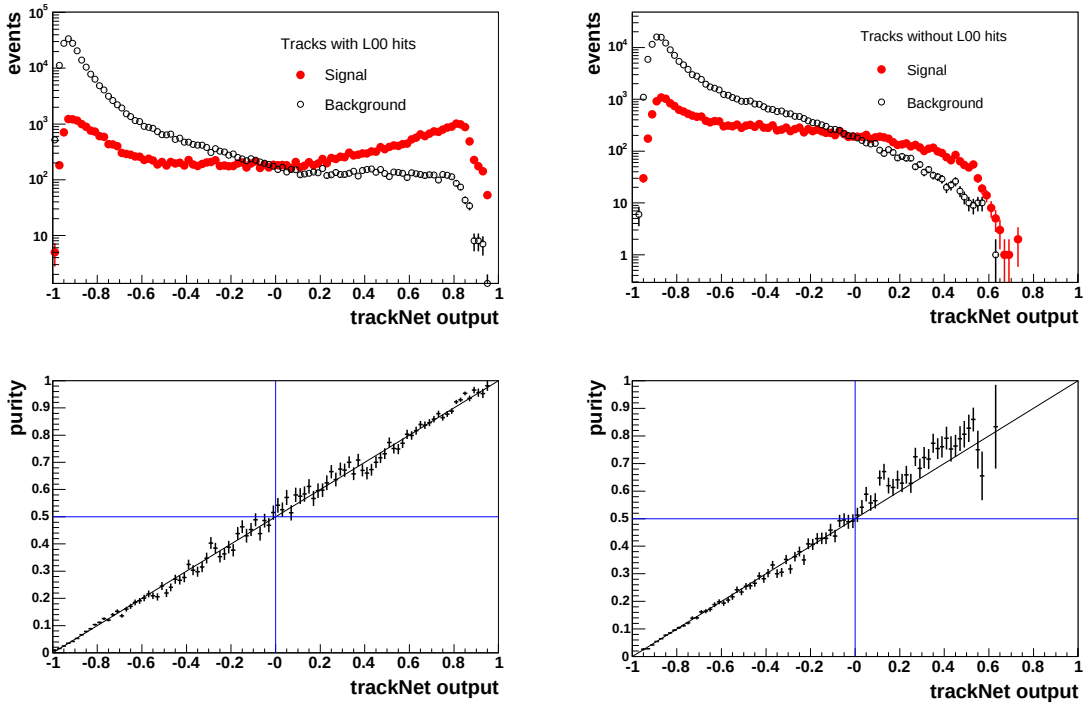

Figure C.2: Result of the training of the trackNet for tracks with L00 hits (*left*) and without (*right*). *Top:* trackNet output for signal and background tracks. *Bottom:* dependence of sample purity on the trackNet output.

the primary or the secondary vertex. The signed impact parameter significance and the magnitude of the impact parameter are correlated and the information about the sign seems to be secondary to the information about the displacement magnitude.
The cone angle between the track and the B candidate, ΔR, has a weak correlation to the target, as expected for a quality variable, nevertheless its removal causes a loss of information of 4σ.
Figure C.2 shows the outcome of the training performed separately for tracks with and without L00 hits. The upper plots illustrates the trackNet output for signal and for background tracks (eqn. 6.1). An excess of signal tracks at high output values is observed and an excess of background tracks is present at low output values.
The lower plots shows the purity of the signal as a function of the output, i.e. the number of signal tracks divided by the number of signal and background tracks in each bin of trackNet output. The purity is distributed along the plot diagonal, indicating a linear dependence on the output. In fact, for a subsample of the tracks in a given bin the purity is p and this means that if a random track is picked from the subsample there is a probability p that the track is signal. In this sense the trackNet output can

signal	d_0^{signed}/σ	track p_T	$\lvert d_0\rvert$	PV	ΔR	p_T^{jet}	y	
	3	2	1	4	6	7	5	sorting
100.0	13.2	13.5	15.1	-4.8	3.0	3.2	5.9	signal
	100.0	2.8	6.7	-2.0	1.1	2.9	1.0	d_0^{signed}/σ
		100.0	-23.5	20.1	0.5	15.7	47.0	track p_T
			100.0	-32.8	3.9	-4.7	-4.8	$\lvert d_0\rvert$
				100.0	-7.9	3.6	-8.1	PV
					100.0	21.2	13.5	ΔR
						100.0	-22.3	p_T^{jet}
							100.0	y

Table C.3: Correlation matrix of input variables after preprocessing for tracks with hits in the silicon detector but no L00 hit. The numbers given are in %.

signal	d_0^{signed}/σ	track p_T	$\lvert d_0\rvert$	PV	ΔR	p_T^{jet}	y	
	4	1	7	5	3	2	6	sorting
100.0	3.6	14.6	-2.0	7.2	5.0	-1.1	8.9	signal
	100.0	1.4	0.2	0.0	0.6	1.6	1.5	d_0^{signed}/σ
		100.0	-17.3	23.9	4.9	25.7	52.1	track p_T
			100.0	-5.9	1.5	-7.2	-6.6	$\lvert d_0\rvert$
				100.0	2.9	-10.3	13.5	PV
					100.0	28.5	7.0	ΔR
						100.0	-12.2	p_T^{jet}
							100.0	y

Table C.4: Correlation matrix of input variables after preprocessing for tracks without silicon hits. The numbers given are in %.

be mapped into the interval $[0, 1]$ and treated as a track probability.

C.2 Separate Training for COT only Tracks

The trackNet for the tagger described in Appendix G has been trained separately for tracks with L00 hits, tracks with hits in the silicon detector but not in L00 and tracks with COT hits only. The training on the first sample has been already described. The training on the second sample has been executed with 23k signal and 161k background tracks; for the third training 1.8k signal and 8.7k background tracks were used, corresponding to all of the tracks with COT hits only assigned to jets in the Monte Carlo sample. The trackNet setup and the input variables have not been changed.
The matrix of correlation between each pair of variables is shown in table C.3 for tracks

with silicon hits and in table C.4 for COT only tracks. Table C.3 is very similar to table C.2, which is relative to all tracks that do not have L00 hits, either COT only or silicon tracks. The order of importance of the variables is essentially the same reported in table C.2, only the last three variables swap places. The importance of variables for the tracks with hits only in the drift chamber is different than in the previous cases: the impact parameter, which used to be the most discriminating variable, is now the least important and the biggest power in separating signal and background is brought by the track p_T, followed by the jet p_T. The resolution on the track displacement for this track is not as good as for the tracks with hits in the silicon detector, consequently the trackNet decision has to rely more on the momentum information.
The result of the training of the trackNet for COT only tracks is shown in fig C.3. It is not possible to reach a sample purity higher than 50% due to the low discrimination power of the input variables. Because of the limited number of tracks available for the training the purity graph shows large fluctuations about the diagonal line.

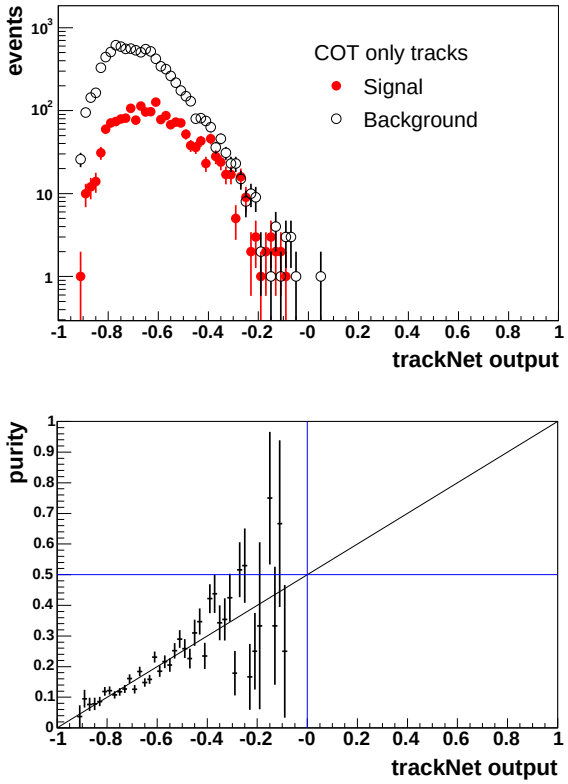

Figure C.3: Result of the training of the trackNet for tracks with hits only in the COT. *Top:* trackNet output for signal and background tracks. *Bottom:* dependence of sample purity on the trackNet output.

Appendix D

Optimisation of the Jet Probability

The bJetNet has 17 input nodes (16 input variables plus the bias node), 8 nodes in the intermediate layer and one output node. A sample of 25k signal jets and 67k background jets has been used for the training. The sample corresponds to all the jets reconstructed by the Cone Clustering algorithm in the ℓ+SVT Monte Carlo sample. The correlation matrix for the input variables is shown in table D.1. The most important variable is the highest track probability for the tracks in the jet t^{max}, followed by the sum of the longitudinal momentum of tracks with respect to the jet axis $\sum_i p_L$ and by the p_T of the jet. The least important variable contributes with a significance of 0.5. Some variables do not bring a very high separation power but during the training NeuroBayes® takes care of killing the intermediate nodes that get a too low weight. The bJetNet output with the breakdown for signal and background and the purity versus output plot are shown in fig. D.1. The same considerations made for the trackNet are valid here, so the bJetNet output scaled to the interval $[0, 1]$ can be interpreted as a probability that the jet is a b-jet.

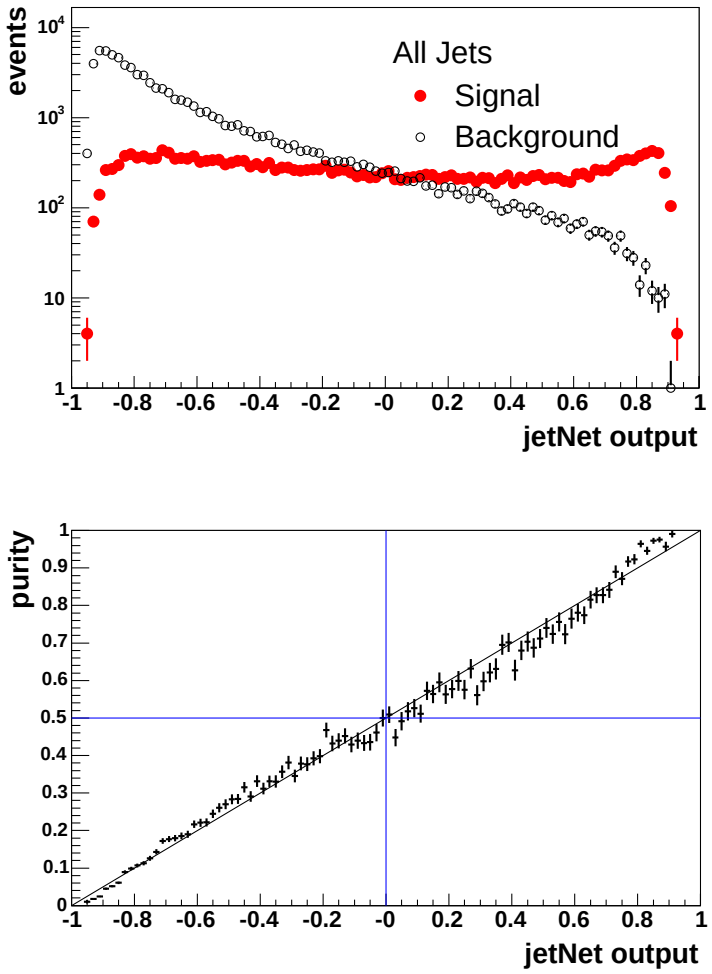

Figure D.1: Result of the training of the bJetNet for jets reconstructed by the Cone Clustering algorithm in the Monte Carlo sample. *Top:* bJetNet output for signal and background jets. *Bottom:* dependence of sample purity on the bJetNet output.

target	N_{50}	p_T^{jet}	N_{SV}	$P(\chi^2_{SV})$	$\frac{\sum cos\alpha}{N_{tracks}}$	$\Delta\phi$	m_{jet}	p_T^{max}	$\ln(J_P^{pT})$	T_{nn}	J_P	N_{tracks}	$\sum p_L$	t^{max}	$\sum_i t_i$	$\sum p_T^{rel}$	
	5	3	4	14	12	8	11	9	6	13	16	15	2	1	10	7	sorting
100.0	45.5	24.3	33.0	29.4	-25.6	20.6	25.1	18.8	20.8	34.3	5.7	24.8	12.4	43.5	42.9	26.1	target
	100.0	34.2	44.0	38.3	-18.1	19.3	31.8	27.5	37.4	55.8	21.8	30.9	31.2	73.9	71.6	31.2	N_{50}
		100.0	25.1	21.1	-54.4	40.1	87.7	79.3	-28.7	61.2	-51.8	83.6	86.8	43.8	69.1	82.7	p_T^{jet}
			100.0	86.3	-13.2	13.3	22.5	22.0	15.0	40.8	4.3	21.6	19.4	33.3	37.1	22.0	N_{SV}
				100.0	-10.4	11.5	18.2	18.8	15.0	34.9	6.3	17.7	16.3	29.0	32.1	17.8	$P(\chi^2_{SV})$
					100.0	-34.5	-77.4	-26.4	25.8	-43.0	56.6	-67.2	-28.6	-24.4	-42.7	-75.9	$\frac{\sum cos\alpha}{N_{tracks}}$
						100.0	42.5	23.3	-2.4	31.7	-21.3	44.8	31.6	27.3	40.8	40.0	$\Delta\phi$
							100.0	55.7	-36.3	66.9	-65.0	91.4	74.4	40.7	68.9	96.8	m_{jet}
								100.0	-8.9	46.4	-20.6	44.7	71.6	35.1	45.8	51.1	p_T^{max}
									100.0	4.8	76.8	-33.1	-14.1	47.3	24.7	-35.5	$\ln(J_P^{pT})$
										100.0	-15.5	61.4	56.1	50.2	69.2	68.5	T_{nn}
											100.0	-65.6	-32.4	25.2	-2.5	-62.8	J_P
												100.0	70.9	39.2	70.0	91.5	N_{tracks}
													100.0	47.8	69.0	68.8	$\sum p_L$
														100.0	88.7	39.4	t^{max}
															100.0	66.6	$\sum_i t_i$
																100.0	$\sum p_T^{rel}$

Table D.1: Correlation matrix for the bJetNet input variables. The numbers given are in %

Appendix E

B-jet Selection Likelihood Variable

A second b-jet selection variable has been developed by combining jet variables with a likelihood ratio method. The input is a subset of the variables used for the jet Neural Network, with the exclusion of the correlated ones. The jets are classified according to the criteria listed in Section 6.3 and signal and background jet are defined in the same way as in Section 6.3. A slightly different approach has been used for the likelihood combination: three input sets have been chosen according to the jet classification and the combination is separate for each set. Some variables have discrimiation power only for a given jet quality or are defined only for a certain class of jets.

E.1 Choice of Input Variables

For each class of jets, an input set has been chosen on the basis of discrimination power between signal and background jets, low correlation among the variables and fairly good data/Monte Carlo agreement. The chosen sets are:

- Class 1
  - jet p_T
  - Secondary vertex fit χ^2 probability
  - number of tracks in the secondary vertex
  - weighted number of tracks in jet $\sum_i t_i$, where the index i runs over all the tracks in the jet and t_i is the trackNet output for track i
  - momentum fraction in the secondary vertex $|\sum_i \vec{P}_i^{SV}|/|\vec{P}_{jet}|$
  - log of jet probability J_P, defined as in 3.2

- Class 2
  - jet p_T
  - highest trackNet output of tracks in jet
  - $\Delta\phi(jet, B)$
  - weighted number of tracks in jet
  - log of $J_P^{p_T}$
  - momentum of the track with the maximum momentum in the jet p_T^{max}
  - momentum fraction in the "secondary vertex" $|\sum_i \vec{P}_i^{SV}|/|\vec{P}_{jet}|$, where the sum is over the tracks that have $t_i > 50\%$
- Class 3
  - jet p_T
  - $\sum_i p_{T,i}^{rel}/p_T^{jet}$, where $p_{T,i}^{rel}$ is the magnitude of the track momentum component perpendicular to the jet axis
  - $\Delta\phi(jet, B)$
  - weighted number of tracks in jet
  - log of $J_P^{p_T}$

Some variables are common to all classes (jet p_T, weighted number of tracks), other are defined only for some classes. Other variables, although defined for every class, have discriminating power only for one jet type ($\sum p_T^{rel}/p_T^{jet}$). The jet probability J_P is computed out of all tracks for Class 1 jets and only with tracks having $p_T > 1$ GeV/c for the other classes because of data/Monte Carlo agreement reasons. For jets of Class 1 the secondary vertex variables are provided by the SecVtx algorithm. For Class 2 jets these variables are computed using the tracks in the jet that have probability greater than 50%. A secondary vertex fit is not attempted in this case, so the χ^2 probability of the fit cannot be computed. For jets in Class 3 the secondary vertex variables cannot be computed.
The correlation between each pair of variables in each set is visualised in figs. E.1 - E.3. These plots have to be looked at like symmetric matrices: each column corresponds to a variable, the n-th column and the n-th row correspond to the same variable. The plots on the diagonal show the fraction of signal in each bin of the variable; the off-diagonal plot in row m and column j is the scatter plot of variable j versus variable m. The off-diagonal plots do not show evident structures, thus indicating that the input variables in each set are weakly correlated with each other.
The comparison of data and Monte Carlo for each variable is shown in figs. E.4 -

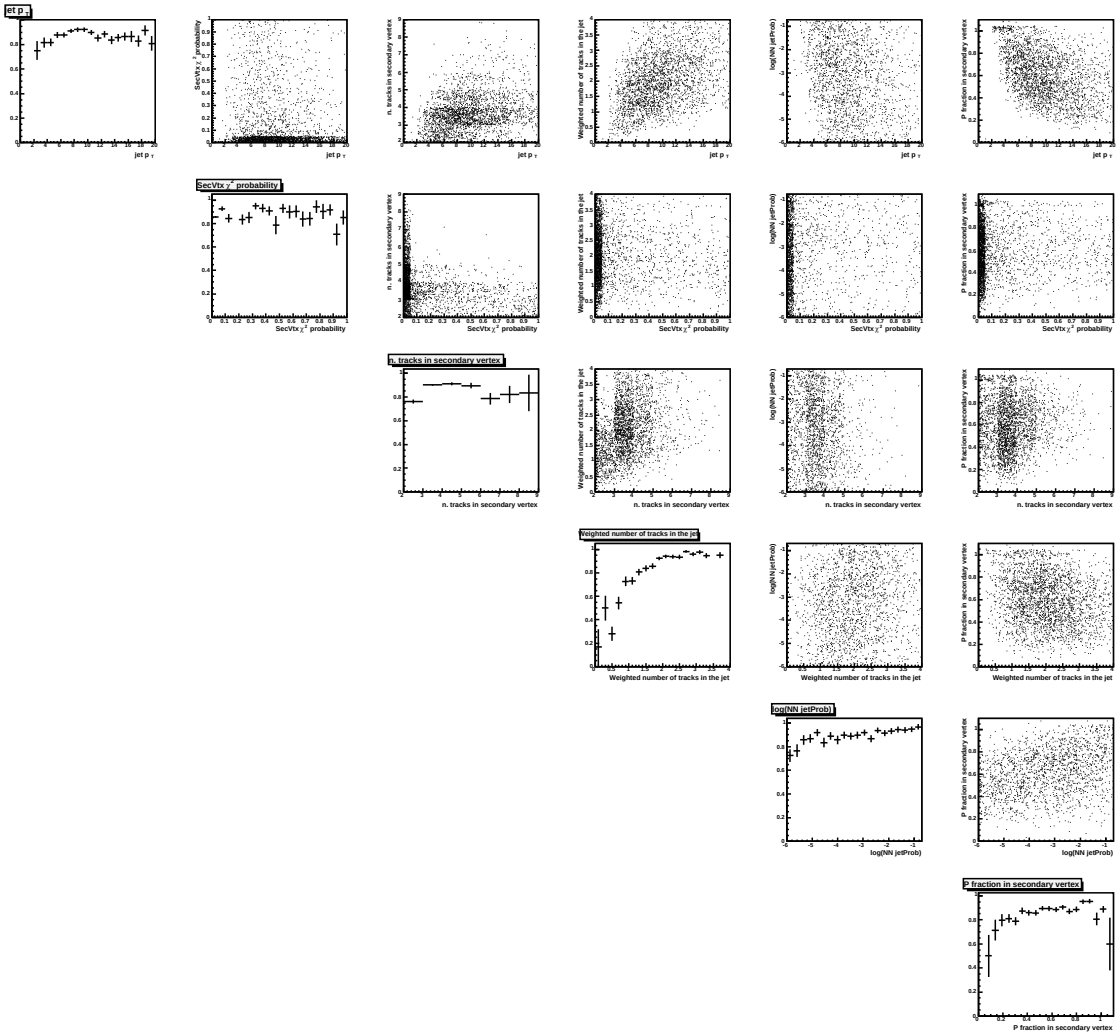

Figure E.1: Correlation plots for variables used for Class 1 jets. Each column (row) corresponds to a variable. The order is the following: jet p_T, secondary vertex fit χ^2 probability, number of tracks in the secondary vertex, weighted number of tracks in jet, log of jet probability, momentum fraction in the secondary vertex. The plots on the diagonal are relative to the signal purity in different bins of the corresponding variable. The off-diagonal plots show the correlation between pairs of variables for signal and background together.

E.6. As already mentioned for the jet Neural Network, the distributions for simulated events do not completely match with the data because of different L00 efficiency in data and in Monte Carlo. An example is given in fig. E.7 of the data/Monte Carlo agreement for some variables when L00 is not included. The use of L00 brings a big improvement to the Jet Charge Tagger although its simulation is not realistic enough.

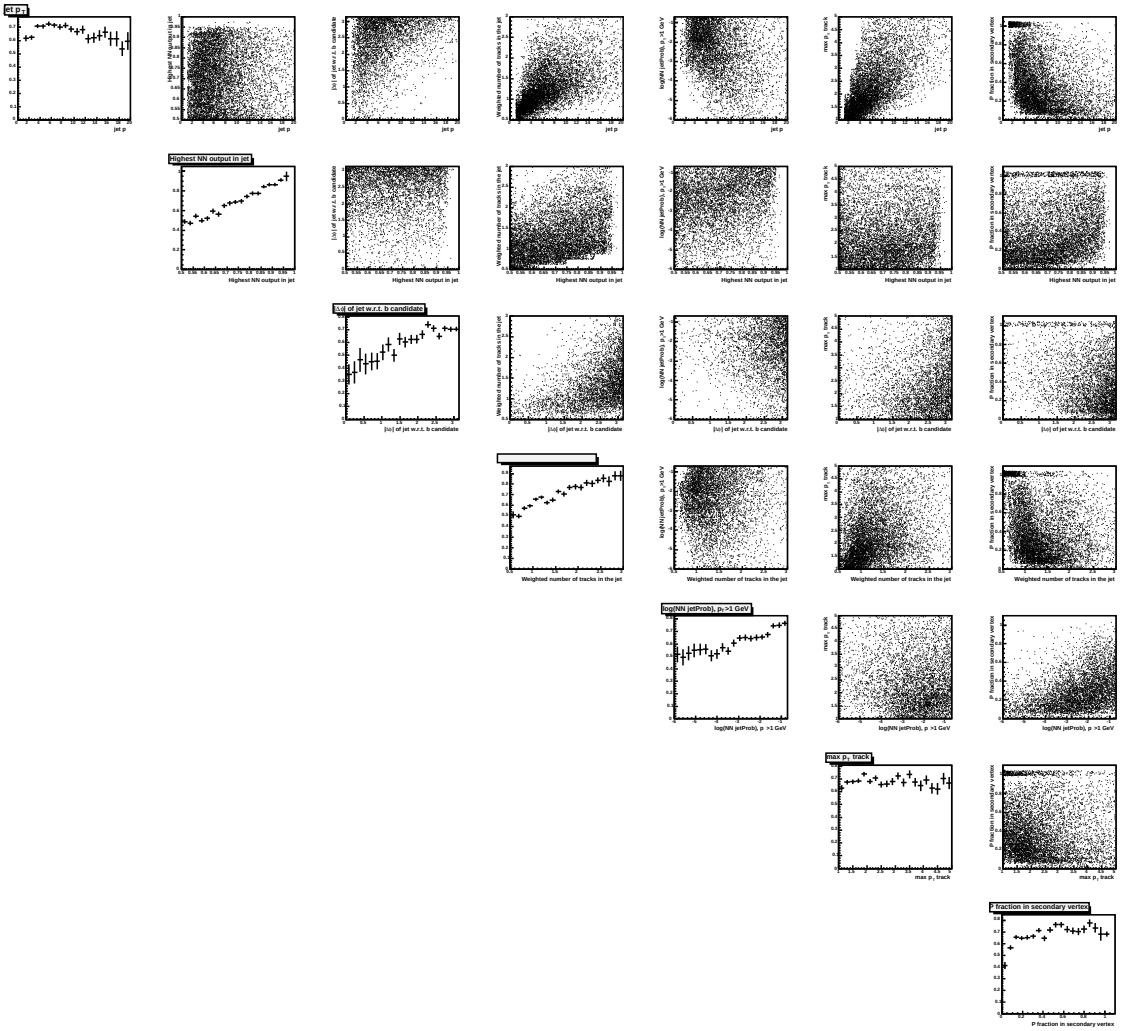

Figure E.2: Correlation plots for variables used for Class 2 jets. The variables for each column/row are: jet p_T, highest probability of tracks in jet, $\Delta\phi(jet, B)$, weighted number of tracks in jet, log of $J_P^{p_t}$, p_T^{max} of tracks in jet, momentum fraction of tracks with probability greater than 50%.

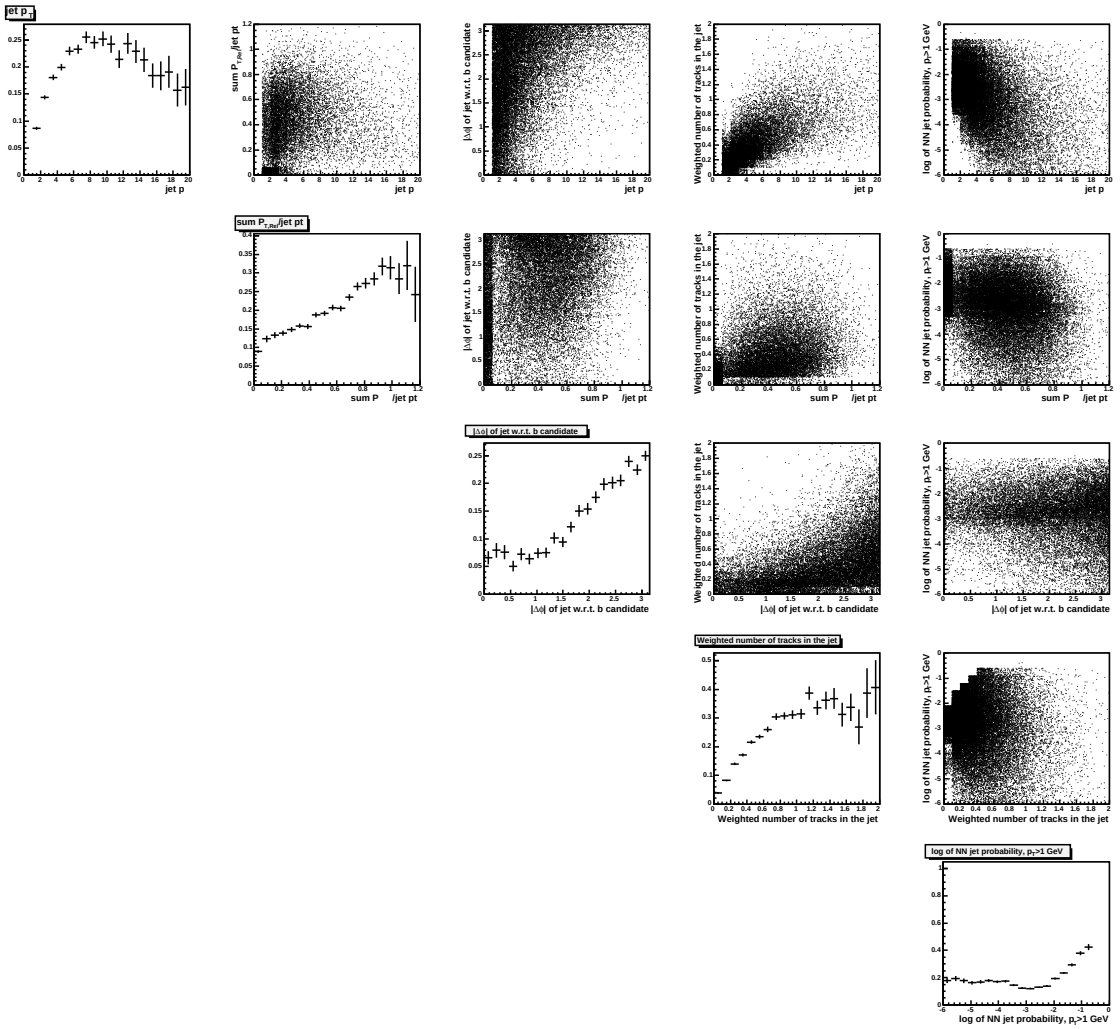

Figure E.3: Correlation plots for variables used for Class 3 jets. The variables corresponding to the columns/rows are respectively: jet p_T, $\sum p_T^{rel}/p_T^{jet}$, $\Delta\phi(jet, B)$, weighted number of tracks in jet, log of $J_P^{p_T}$.

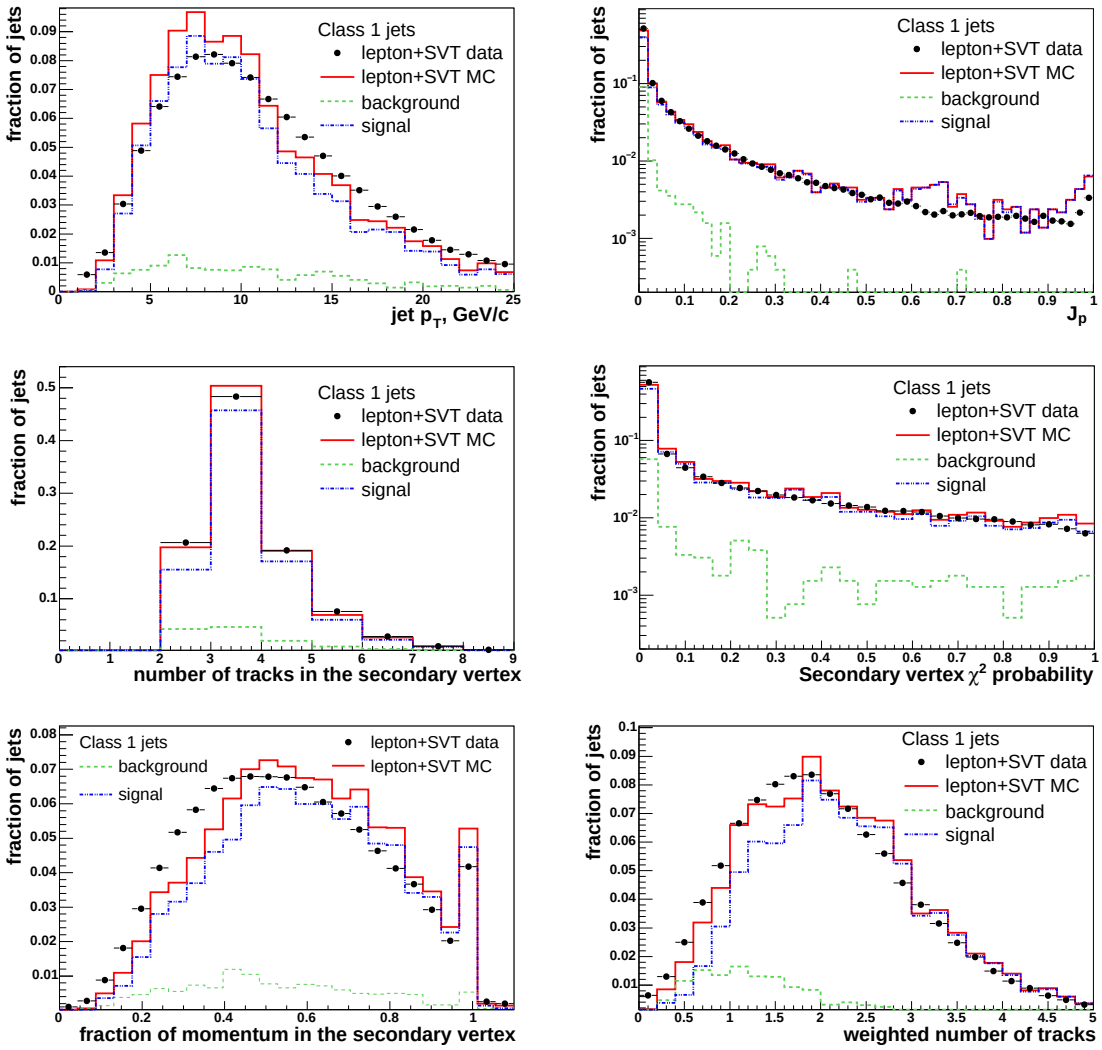

Figure E.4: Input variables for Class 1 jets: jet transverse momentum, J_P, number of tracks in the secondary vertex, probability of SecVtx χ^2 fit, fraction of the jet momentum in the secondary vertex, weighted number of tracks.

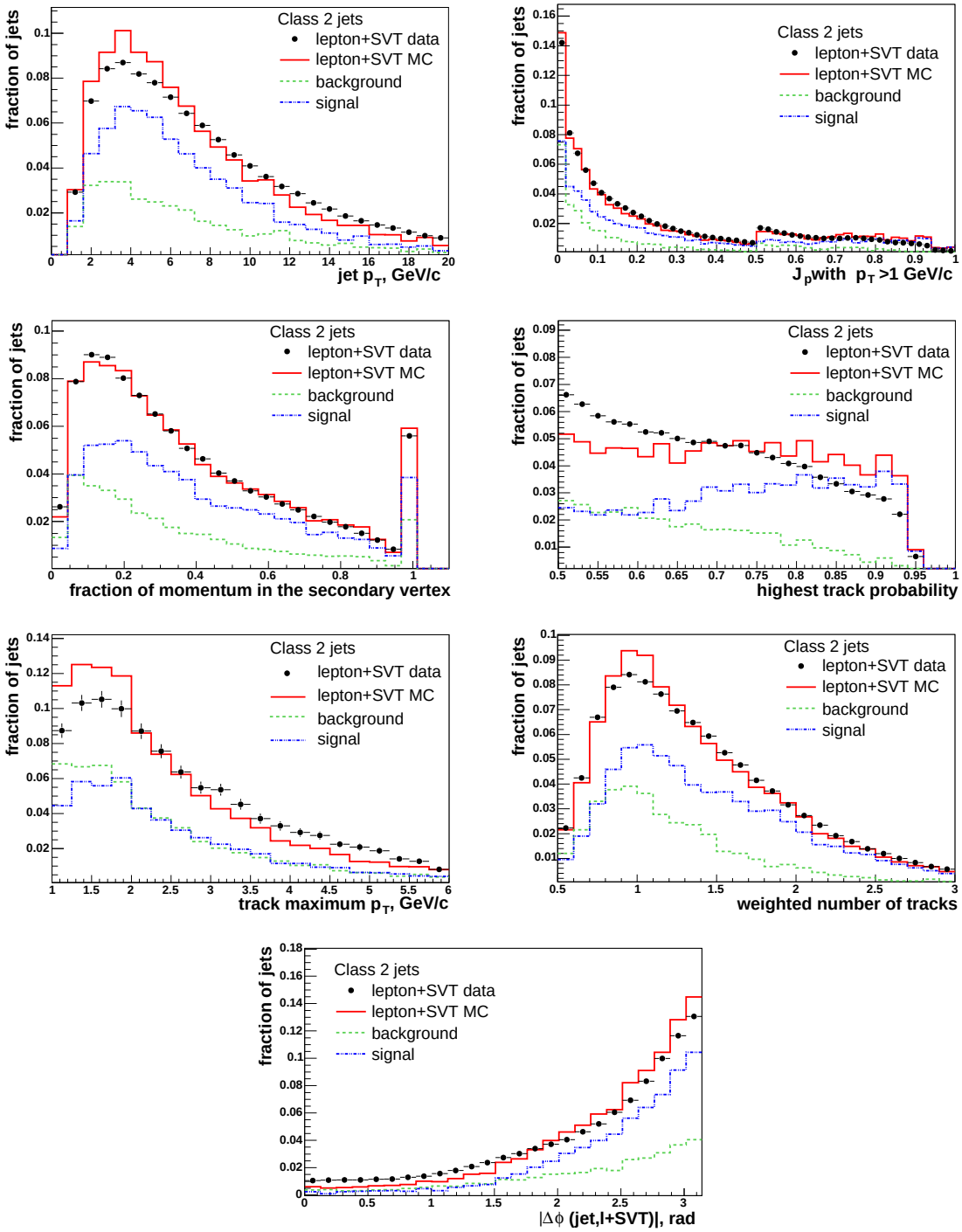

Figure E.5: Input variables for Class 2 jets: jet p_T, highest probability of tracks in jet, $\Delta\phi(jet, B)$, weighted number of tracks in jet, log of J_P^{pt}, p_T^{max} of tracks in jet, momentum fraction of tracks with probability greater than 50%.

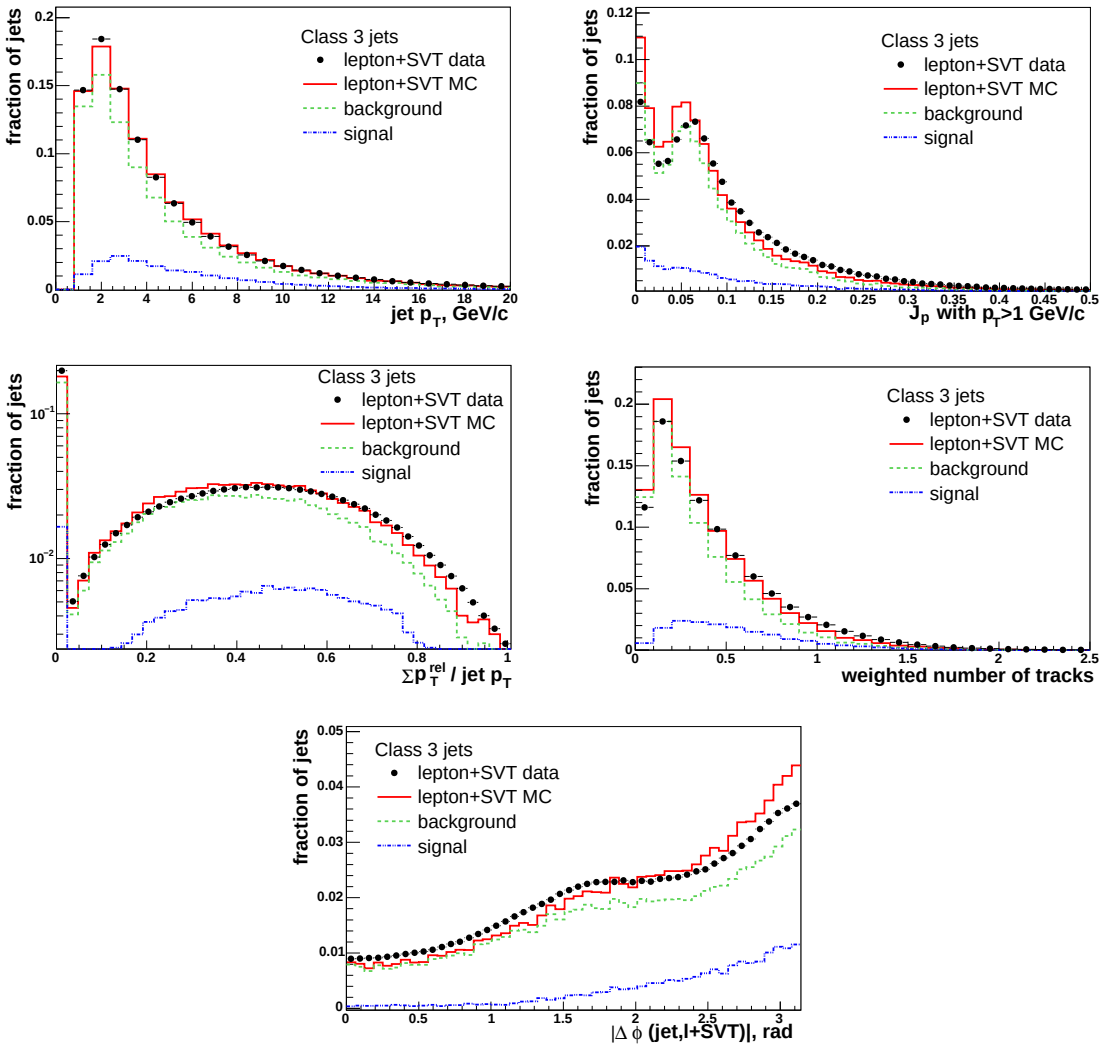

Figure E.6: Input variables for Class 3 jets: $\sum p_T^{rel}/p_T^{jet}$, $\Delta\phi(jet, B)$, weighted number of tracks in jet, log of $J_P^{p_T}$

E.2 Likelihood Ratio

For each input variable in each set, the Probability Density Function (PDF) for the signal, f^S, and for the background, f^B, is extracted from the ℓ+SVT Monte Carlo. For a variable that takes the value x, $f^S(x)$ is the probability that the jet is signal and $f^B(x)$ is the probability that the jet is background.
The jet likelihood variable is computed by combining the input variables for the subset

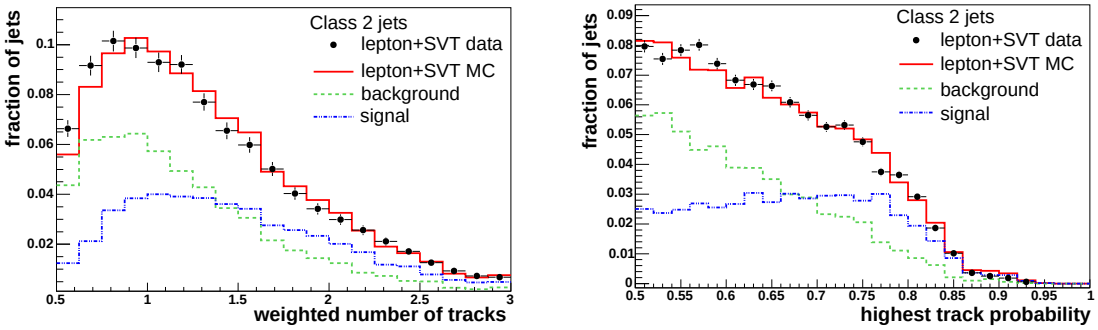

Figure E.7: Distributions of the weighted number of tracks and the highest trackNet output for Class 2 jets when the L00 is not used. The data/Monte Carlo agreement is significantly better than in the case in which L00 hits are added to the tracks.

corresponding to the jet class with the formula

$$L = -\ln\left(\prod_i \frac{f_i^B(x_i)}{f_i^S(x_i)}\right) \tag{E.1}$$

where the index i runs over the variables in the subset. Possible correlations among variables are not taken into account in eqn. E.1.
For a signal jet the background probability should be lower than the signal probability for some or all of the variables. The probability ratio should be then smaller than 1 and the negative log in formula E.1 gives a positive likelihood. The likelihood value is expected to be negative for a background jet.
The development of a multi-dimensional likelihood variable has not been considered, due to lack of statistics, especially for Class 1 jets.

E.3 Performance

The distribution of the likelihood ratio for data and simulated events is shown in fig. E.8. It is shifted towards higher likelihood values for signal jets and towards low likelihood values for background jets.
The intervals in which the likelihood is distributed are not the same for the three classes because of the different number of variables that are combined for each set, therefore the likelihood of jets in different classes cannot be directly compared. This has to be taken into account when the jet selection is based upon the likelihood variable.
The performance of the likelihood variable for the three classes of jets is shown in fig. 6.24 (right) in comparison with the bJetNet output. The sample purity increases

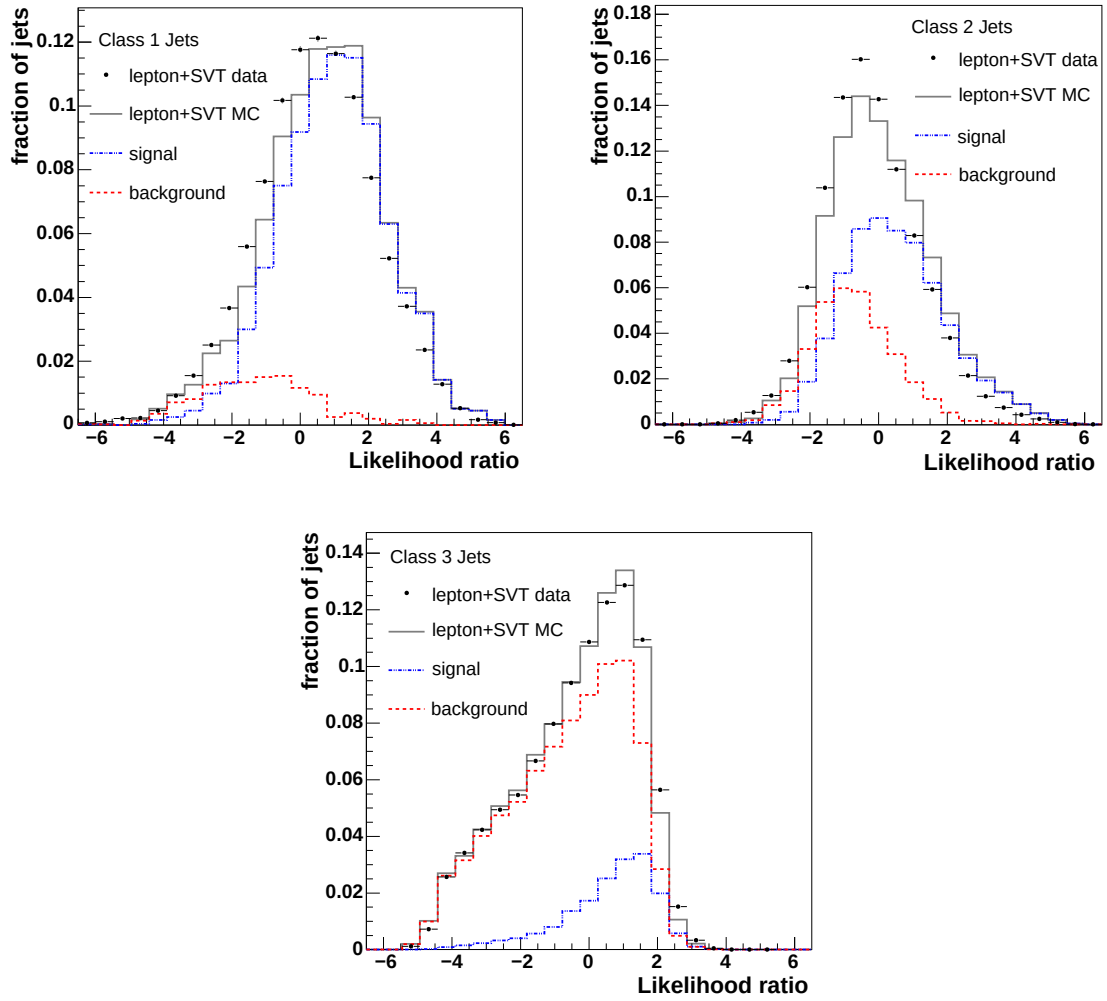

Figure E.8: Likelihood for Class 1 (*left*), Class 2 (*center*) and Class 3 jets (*right*) for Monte Carlo events. The signal and background component are shown (respectively dotted and dashed line) as well as the distribution for the data (dots).

as the cut becomes larger. The growth varies according to the jet class. For Class 1 jets the purity reaches a plateau after removing jets with the lowest likelihood. Class 2 shows an almost linear dependence of purity on the likelihood cut. A pure sample of Class 3 jets is selected only with a hard cut. As already mentioned in Section 6.3.2, the likelihood does not perform as well as the Neural Network because the correlated variables, which bring additional power, have been excluded from the input set.

Appendix F

Likelihood based Jet Charge Tagger

An alternative Jet Charge Tagger basically identical to the one described in Chapter 7 except for the jet selection variable was developed. The likelihood variable described in Appendix E is exploited as a jet selection variable. The jet selection on the opposite side is implemented in the following way:

- a Class 1 jet is chosen
- if no Class 1 jet exists in the event, a Class 2 jet is chosen. In case more than one such jet exists, the one with the highest likelihood is selected
- if no Class 1 and Class 2 jets exist, the jet with highest likelihood in the event is the tagging jet

The construction of the likelihood variable makes this apparently complicated selection strategy necessary. Since the likelihood value of jets of different classes cannot be directly compared, the higher purity of Class 1 and Class 2 jets helps in setting selection priorities. The likelihood distribution for tagging jets is shown in fig. F.1. A shift towards high values with respect to the likelihood distributions for tagging and non-tagging jets (fig. E.8) is observed.
No requirements are applied to the tagging jet. The tagging power is computed following the method described in Section 7.3. The jet sample is split in 10 bins of $|Q_{jet}| \cdot L$, where L is the likelihood variable scaled to the interval $[0, 1]$. The result of the evaluation is shown in table F.1. The measured efficiency is the same as of NNJQT (table 7.3) because the sample of jets is the same in both cases and no cuts are applied to the tagging jet. The effective dilution is $\sim$0.5% lower, averaging the result on the e+SVT and the μ+SVT samples, because of the slightly worse performance of

Jet type	ϵ,%	effective D, %	ϵD^2
e+SVT data			
Class 1	11.05 ± 0.05	18.46 ± 0.45	0.377 ± 0.018
Class 2	28.92 ± 0.08	11.20 ± 0.32	0.363 ± 0.020
Class 3	55.50 ± 0.12	4.86 ± 0.21	0.131 ± 0.011
combined	95.48 ± 0.15	9.55 ± 0.16	0.870 ± 0.030
μ+SVT data			
Class 1	11.04 ± 0.04	17.34 ± 0.39	0.332 ± 0.015
Class 2	29.52 ± 0.07	10.77 ± 0.27	0.343 ± 0.017
Class 3	55.17 ± 0.09	5.92 ± 0.19	0.134 ± 0.010
combined	95.72 ± 0.12	9.19 ± 0.14	0.808 ± 0.025

Table F.1: Summary of ϵ, D and ϵD^2 measured on data for the Jet Charge Tagger based on likelihood jet selection.

the likelihood variable (see fig. 6.24). The result is anyway remarkable since it is in average ∼12% better than the cut based Jet Charge Tagger.

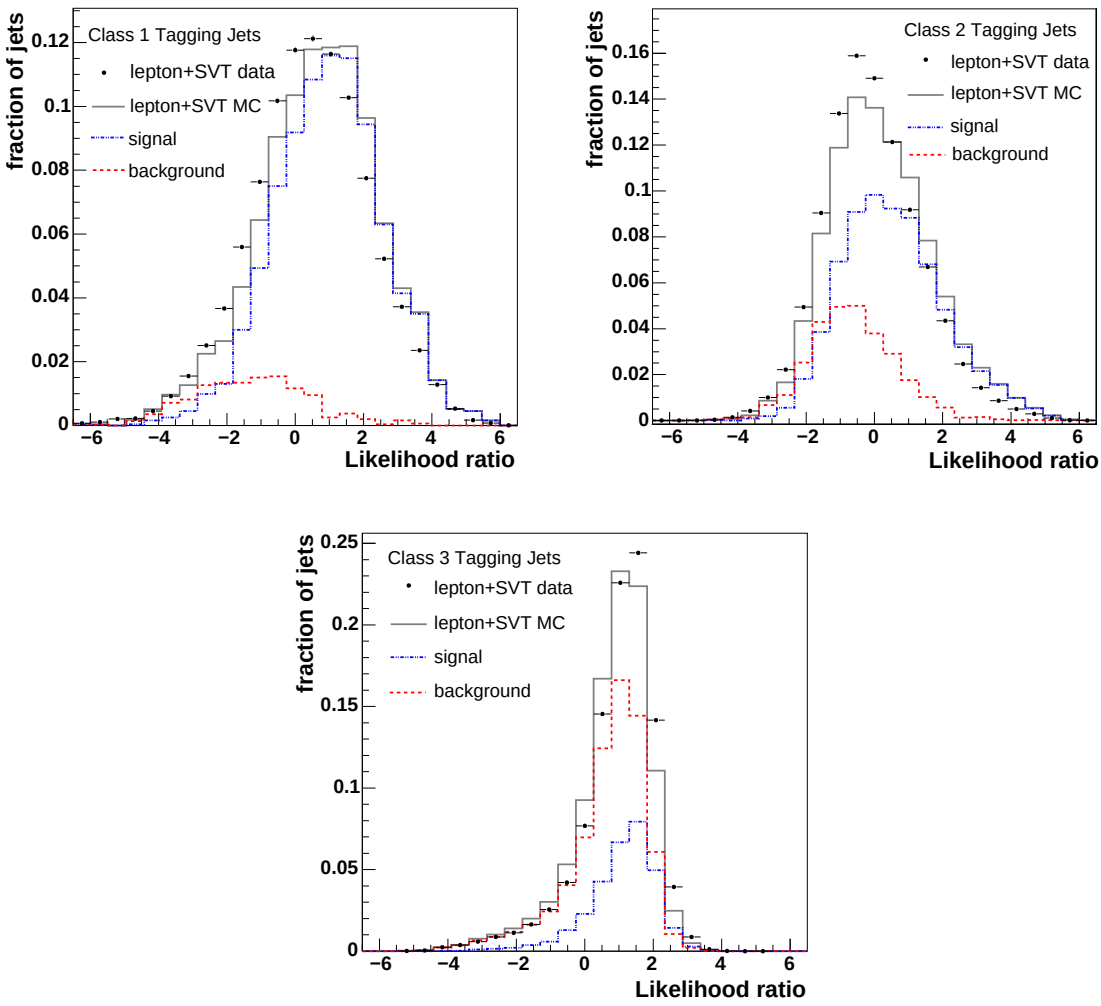

Figure F.1: Likelihood for tagging jets in Class 1 (*left*), Class 2 (*center*) and Class 3 (*right*). The signal and background components are shown (respectively dotted and dashed line). The dots represent the data distribution.

Appendix G

Further Development of the Jet Charge Tagger

To take advantage of the better resolution on the impact parameter of tracks with L00 hits, the track probability has been trained separately on tracks with and without L00 hits. Since tracks without hits in the silicon detector have different impact parameter resolution than tracks with silicon hits, a gain in the performance of the track probability Neural Network could come from a separate training on tracks with hits only in the COT. Of all the tracks assigned to jets, only 2.7% do not have silicon hits. The fraction of these tracks in the event before the jet reconstruction is performed is $\sim$14% and it is reduced by the d_0 and Δz_0 requirements on the tracks used for the clustering (see Section 6.1.3). The distributions of d_0 and Δz_0 for all tracks and for COT only tracks are in fig. G.1, where the arrows indicate the position of the cuts applied before jet clustering. The cuts were chosen after an optimisation on ϵD^2 carried on at an earlier stage of the cut based Jet Charge Tagger development and described in [47].
This Appendix describes the result obtained after training the trackNet for three track samples: tracks with L00 hits, tracks with silicon hits but no L00 hits, tracks with hits only in the COT. The result is propagated to the jet Neural Network and to the Neural Network based Jet Charge Tagger.
An alternative solution to the splitting of the track sample into three samples would be to introduce the number of silicon hits in the list of input variables for the trackNet. As it has been shown in Section 5.4, the modelling of this variable in the Monte Carlo is not satisfactory, therefore the alternative solution has been rejected.

G.1 Track Probability Optimisation

The training for tracks with L00 hits is the same described in Section 6.2. The details about the trackNet training with tracks with silicon hits and without L00 hits, and

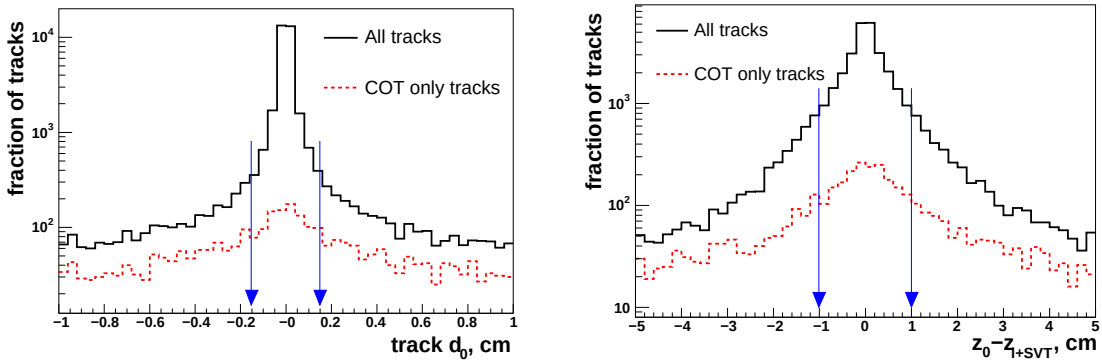

Figure G.1: Distribution of d_0 (*left*) and Δz_0 (*right*) for all tracks and for tracks with hits only in the drift chamber. The arrows show the position of the cut that the tracks have to satisfy to be accepted for jet clustering.

with COT only tracks are in Appendix C.2. The track probability for COT only tracks obtained with this training is shown in fig. G.2. The performance of the separate training is shown in fig G.3 (left). The purity reached by COT only tracks is about half of the purity that tracks with hits in the silicon detector can achieve, because of the discriminating power of variables like the impact parameter, which is higher in the second case.
The improvement in the b-jet Neural Network performance caused by the new trackNet training is illustrated in fig G.3 (right), where it is noticeable that the new bJetNet (dashed line) performs slightly better than the standard (solid line) at high efficiency.

In order to understand how important the d_0 information is in determining the track probability for each class of tracks, the contribution to the network performance of $|d_0|$ and d_0/σ is shown in fig. G.4 for different training. The curves shown for $|d_0|$ and d_0/σ can be interpreted as the performance of a Neural Network that has only that variable as input. When the track probability is trained with silicon and COT tracks together (fig. G.4, bottom right) the power brought by the displacement information is basically the same as when the COT tracks are removed (fig. G.4, top right). The COT only tracks are not significantly diluting the power of $|d_0|$ and d_0/σ in the inclusive sample, as it is expected considering the small fraction of these tracks. Therefore the training on three track samples is not necessary.

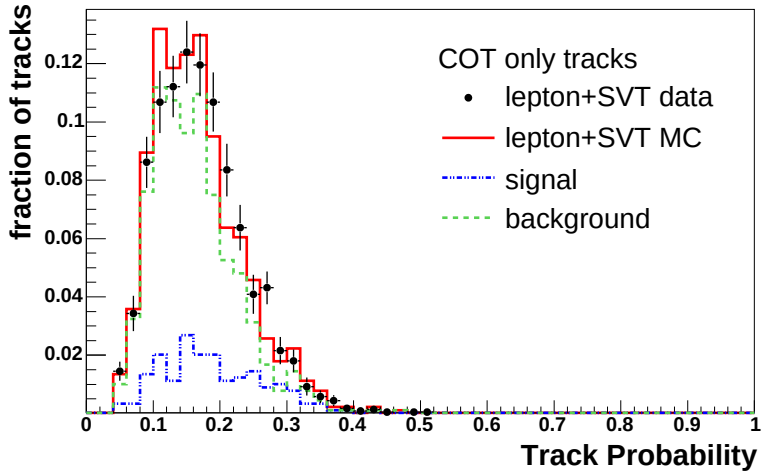

Figure G.2: Distribution of the trackNet output for tracks without silicon hits.

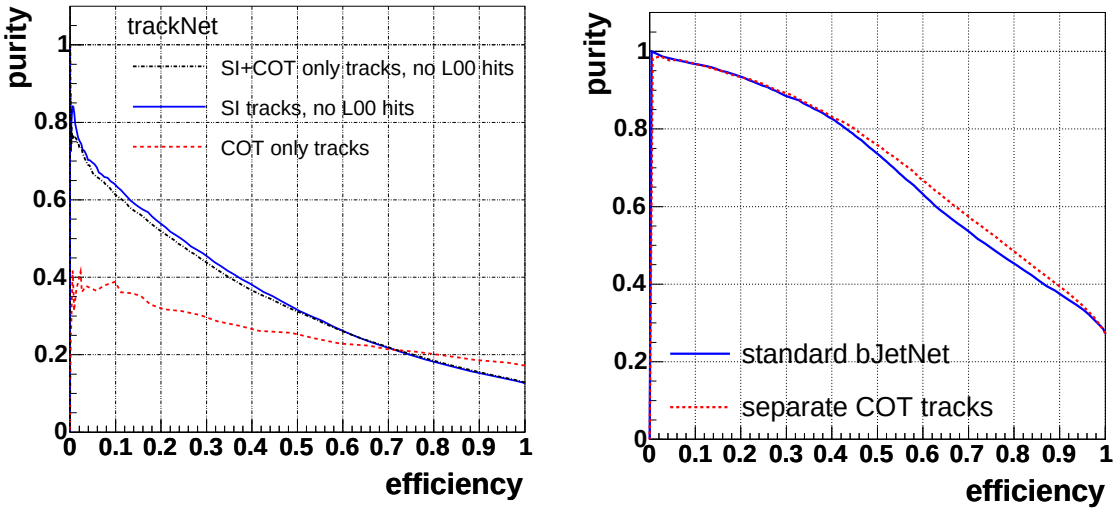

Figure G.3: *Left:* Performance plot of the trackNet when the training on COT tracks is performed apart of tracks with hits in the silicon detector. The dash-dotted line represents the performance of the training for tracks without L00 hits, considering silicon and COT tracks together *Right:* Performance of the bJetNet that uses the trackNet trained separately on three track samples (*dashed line*) compared to the bJetNet described in Section 6.3 (*solid line*)

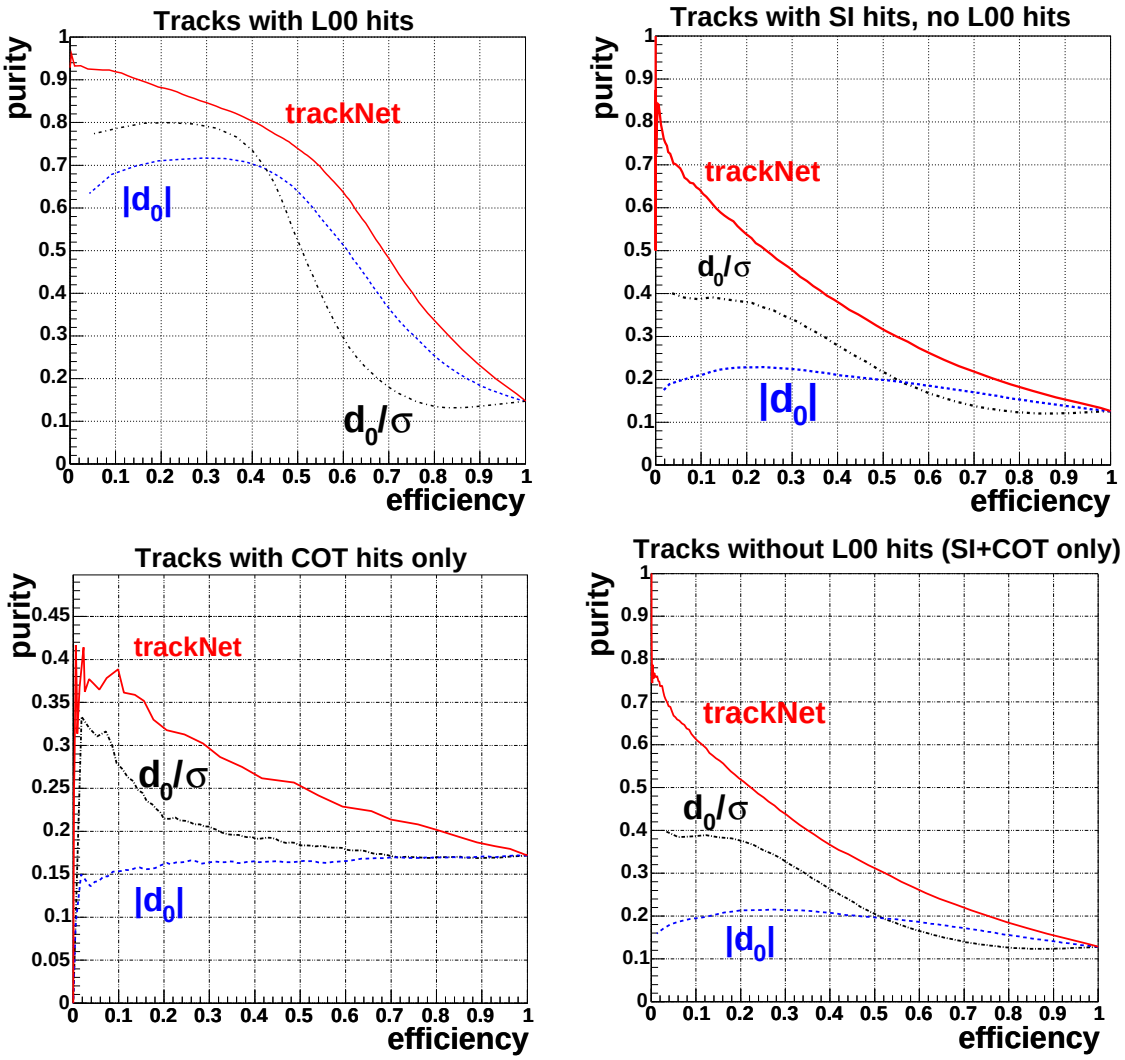

Figure G.4: Contribution of $|d_0|$ (*dashed line*) and d_0/σ (*dash-dotted line*) to the trackNet performance (*solid line*) for the training on different track samples. *Top left:* tracks for which a L00 hit was found *Top right:* tracks with hits in the silicon detector, without L00 hits *Bottom left:* tracks with hits only in the COT (notice the different y-axis scale) *Bottom right:* tracks without L00 hits, either with or without hits in the silicon detector.

Jet type	ϵ,%	effective D, %	ϵD^2
e+SVT data			
Class 1	10.77± 0.05	19.66± 0.41	0.416± 0.017
Class 2	28.24± 0.08	11.48± 0.34	0.390± 0.022
Class 3	58.47± 0.11	4.76± 0.35	0.128± 0.019
combined	95.48± 0.15	9.89± 0.17	0.934± 0.034
μ+SVT data			
Class 1	10.73± 0.04	17.88± 0.35	0.343± 0.013
Class 2	28.81± 0.07	12.23± 0.30	0.431± 0.021
Class 3	56.19± 0.09	5.60± 0.32	0.176± 0.020
combined	95.72± 0.12	9.96± 0.15	0.950± 0.029

Table G.1: Tagging power of NNJQT measured on the ℓ+SVT data sample. The track probability Neural Network has been trained independently on three track samples (tracks with L00 hits, tracks with hits in the silicon detector but no L00 hits, tracks with hits only in the COT). The numbers listed here have to be compared to those in table 7.3.

G.2 Tagger Performance

The NNJQT has been run using the new track and jet probability. The setup has not been changed with respect to the description given in Chapter 7. The performance has been evaluated in the same way and the results are listed in table G.1. The tagging power has increased of $\sim$1.5% with respect to the results shown in table 7.3. The variation is smaller than the error on the measurement.

Appendix H

Tagger Usage Instructions

Before compiling any code depending on the NeuroBayes® package [70], it is necessary to setup the `neurobayes` product

```
setup neurobayes v1_0 -q KCC_4_0
```

The environmental variable `NEUROBAYES_DIR` is set to point to the directory where the libraries and licence are installed[1]. The product provides the Expert class, which is the C++ interface to NeuroBayes® expertise files.

H.1 The TrackProbNN Class

The interface to call the track probability neural network is implemented in the TrackProbNN class, in `BottomTaggers/BottomTaggers/TrackProbNN.hh` and belongs to the BottomTaggers package [72]. The track probability can be used with one, two or three track categories:

1 category All tracks

2 categories
- tracks with L00 hits
- tracks without L00 hits

3 categories
- tracks with L00 hits
- tracks without L00 hits and with at least one hit in the silicon detector
- tracks with hits only in the COT

[1]The `neurobayes` product is available at Fermilab and at the Institut für Experimentelle Kernphysik in Karlsruhe. In order to obtain a licence for the product it is necessary to contact <phi-t> (licence@phi-t.de). More information about NeuroBayes® can be found on the web: http://www.phi-t.de.

by setting a different number of expertises with the methods

```
setExpert(Expert* net1)
setSecondExpert(Expert* net2)
setThirdExpert(Expert* net3)
```

By default two categories are used. The expertises prepared for the results shown in Section 6.2 are in the directory `BottomTaggers/constants` and are `pbnet_cone_sa_all_l00.nb` for tracks with L00 hits and `pbnet_cone_sa_all_nol00.nb` for tracks without L00.
The TrackProbNN constructor needs a pointer to the event. In order to get a meaningful output for a track, it is necessary that besides the expertise(s) the direction of the reconstructed b-hadron, the jet to which the track belongs and the beam line are set. This is achieved by calling, respectively, the methods `setTrigger(Hep3Vector p)`, `setJet(ProtoJet* jet)` and `setBeamline (Beamline beam)`.
The method `getTrackProbability(const CdfTrack* track)` computes the input variables, calls the trackNet expertise(s) and returns a value between 0 and 1. The return value is a dummy (999.) in case the beam line is invalid or the neurobayes package is not installed on the node where the class is compiled.

H.1.1 How to Use TrackProbNN in AC++

In order to use the TrackProbNN class in an AC++ module, it is necessary that the header of this class and of the Expert class are included:

```
#ifdef NEUROBAYES_DIR
#include "Expert.hh"
#endif
#include "BottomTaggers/TrackProbNN.hh"
```

The lines of code for the initialisation are

```
 AbsEvent* event = AbsEnv::instance()->theEvent();
 TrackProbNN* tProb = new TrackProbNN(event);
 // set the direction of the reconstructed B
 Hep3Vector bDirection = ...;
 tProb->setTrigger(bDirection);
 // set the beamline
 Beamline beam = ...;
 tProb->setBeamline(beamline);
 // set the expertise(s)
#ifdef NEUROBAYES_DIR
```

```
 // load the expertises
 Expert* trackNet = new Expert("pbnet_cone_sa_all_l00.nb");
 Expert* trackNet2 = new Expert("pbnet_cone_sa_all_nol00.nb");
 // set the expertises
 tProb->setExpert(trackNet);
 if (trackNet2!=0) tProb->setSecondExpert(trackNet2);
#endif
```

The path specified for the expertise files has to correspond to the location of the files at running time. The probability for a single track is obtained with the following lines of code:

```
 CdfTrack_clnk theTrack = ...;
 // set the jet to which the track belongs
 ProtoJet* jet = ...;
 tProb->setJet(jet);
 // get the probability
#ifdef NEUROBAYES_DIR
 double trackProbability = tProb->getTrackProbability(&*theTrack);
#endif
```

The objects declared as `new` should be properly deleted at the end of the program.

H.2 The NNJetQTagger Class

The Neural Network based Jet Charge Tagger is implemented in the NNJetQTagger class, in the BottomTaggers package. The header file is `BottomTaggers/BottomTaggers/NNJetQTagger.hh`. This class inherits from the abstract class AbsTagger (`BottomTaggers/BottomTaggers/AbsTagger.hh`) and uses the TrackProbNN class. The interface to the jet probability neural network is not a separate class but a method of the NNJetQTagger class.
The selection of tracks on the opposite side is performed in the `fillNonBTracks()` method. The main method is `tag()`. At first the clustering algorithm is run. The default choice is Cone Clustering, but Mass Clustering can be run by setting the appropriate switch. Then the method `JetSelectionAlg::selectBySvtx(SecVtxAlg* extSecVtxAlg)`, which interfaces SecVtxAlg, is called to search for a secondary vertex among the jets. The jet probability method `bJetNetOutput()` is called in a loop over the jets to compute the jet variables and to call the expertise. The jet probability is a double variable with values distributed between 0 and 1. If the code is not correctly compiled with neurobayes, then the probability variable has a dummy value.

The expertise employed to produce the results in Section 7.3 can be found in `BottomTaggers/constants/jetnet_cone_sa_all_l00.nb`.
The tagger decision can be accessed via the method `getDecision()`, which returns the opposite of the sign of the Q_{jet} variable. The expected dilution is computed with the parameters in table 7.2 (no SLT exclusion) and is returned by the method `getWeight()`.

H.2.1 How to Use NNJetQTagger in AC++

The following lines have to be added to the header file of the module in which NNJetQ-Tagger should be called

```
#include "BottomTaggers/NNJetQTagger.hh"
#include "TrackingSI/Utils/SiExpected.hh"
```

and among the module private data members these objects should be added

```
SiExpected          _siExpected;
NNJetQTagger*       _jqTagger;
```

The source file of the module needs to be modified as follows. In the `beginJob()` method the lines

```
_jqTagger->setTrackNet("pbnet_cone_sa_all_l00.nb");
_jqTagger->setTrackNet2("pbnet_cone_sa_all_nol00.nb");
_jqTagger->setJetNetFile("jetnet_cone_sa_all_l00.nb");
_jqTagger->init();
```

have to be included. In the `beginRun()` method the silicon maps are loaded via the command:

```
bool status = _siExpected.beginRun();
```

The tagging itself is executed in the `event()` (or `fillHistogram()`) method. Several elements are needed for the tagger to work correctly

- an event pointer, which comes into the module which calls the tagger as an argument of the event method. Below it is called "anEvent".
- the list of all tracks that should be used by the tagger, which will automatically exclude tracks that belong to the signal B candidate and tracks in cone around the signal B within $\Delta R = 0.7$. It will also remove tracks with impact parameter larger than 0.15 cm and those that fail the cuts on the same side B daughter rejection. A possibility is to use the list of defTracks, or the list of tracks refitted

by the user. All kind of tracks might be in the list, including SISA and IO. In order to reproduce the blessed ϵD^2 numbers, the tracks have to be refitted. The refit should be performed adding L00 hits and using the blessed settings for the covariance matrix scaling and energy loss corrections documented in [57]. An object of the type CdfTrackView_ch has to be passed to the tagger (below it is called "tracks").

- the list of all tracks used to reconstruct the B-candidate in an object of type CdfTrackView_ch ("btracks").
- a Hep3Vector holding the B-candidate direction ("bDirection") and one holding the position of the B-candidate decay vertex("bDecayVertexPos").
- a double variable ("zPrimaryVertex") may be filled by the true primary vertex z-coordinate, or just with the z-position of the B-candidate vertex, or with an average of the z_0 for all tracks used to reconstruct the B-candidate. In all cases the precision will be about $\sim$1 mm, which is enough for Jet Charge tagging purposes.
- the primary vertex is the beam line position at the z-coordinate of primary vertex. This is not an accident. The NNJQT has been calibrated with the use of the beam line r-ϕ position. The z-coordinate of the event-by-event primary vertex might be used, but still the beam line in r-ϕ should be used as in the example below to keep the predicted dilutions valid.
- the beam line has to be passed as well. An object of type SvxBeam should be present in the module and properly loaded. Below this object is called "svxBeam".

All these objects are passed as follows

```
_jqTagger->clear();

// Fill information for the tagger to work with

_jqTagger->setEvent(Event);
_jqTagger->setZPrimaryVertex(zPrimaryVertex);
_jqTagger->setPrimaryVertex(svxBeam->position(zPrimaryVertex));
_jqTagger->setPrimaryVertexCov(svxBeam->cov3(zPrimaryVertex));
_jqTagger->setBeamline(svxBeam->getBeamline());
_jqTagger->setTracks(tracks);
_jqTagger->setBCandidate(btracks, bDirection, bDecayVertexPos);
```

```
_jqTagger->setSiExpected(&_siExpected);

// Do the tagging

_jqTagger->tag();
```

The tagger decision and the expected dilution are retrieved as it is shown below:

```
double decision = _jqTagger->getDecision(); // 0 or -999 mean no tag
int    jetType  = _jqTagger->getTagType();  // Class1/2/3 NN jet
double Qjet     = _jqTagger->getJetCharge();// -999 means no tagging jet
double Dpred    = _jqTagger->getWeight();   // predicted true dilution

// Ntupling can be done here
...
```

H.3 How to Compile BottomTaggers

To compile and link the code in BottomTaggers correctly, and reproduce the blessed results, several packages need to be checked out. Here it is assumed that the user is working with the cdfsoft2 release 5.3.4.

```
addpkg -h BottomTaggers
addpkg -h JetUserObjects
addpkg -h BottomMods
addpkg BTagAlgs         V00-00-19
addpkg BTagObjects      V00-00-44
addpkg BTagMods         V00-00-43
```

Three classes need to be patched

```
patch -R BTagAlgs/src/SecVtxTrackSelector.cc \
 < BottomMods/PATCH/5.3.3_SecVtxTrackSelector.cc.diff
patch -R BTagAlgs/BTagAlgs/SecVtxAlg.hh \
 < BottomMods/PATCH/5.3.4_SecVtxAlg.hh.diff
patch -R BTagAlgs/src/SecVtxAlg.cc \
 < BottomMods/PATCH/5.3.4_SecVtxAlg.cc.diff
patch -R BTagObjects/BTagObjects/SecVtxParams.hh \
 < BottomMods/PATCH/5.3.4_SecVtxParams.hh.diff
```

and after this action is taken, the BottomMods package can be removed

```
rmpkg BottomMods
```

The libraries can be compiled after the neurobayes product has been set up

```
setup neurobayes v1_0 -q KCC_4_0
```

This command has to be executed before the setup of cdfsoft2 because neurobayes might set a different root version than the one set by cdfsoft2.

```
gmake BottomTaggers.nobin
gmake JetUserObjects.nobin
gmake BTagAlgs.nobin
gmake BTagObjects.nobin
gmake BTagMods.nobin
```

The file `test/GNUmakefile` of the package to which the module belongs has to include the dependency on BottomTaggers (e.g. `YourPackageName`). Before the list of LINK entries, these lines should be present:

```
override LOADLIBES += -lYourPackageName -lBottomTaggers \
                      -lJetUserObjects
```

In the top section of the LINK list:

```
override LINK_BottomTaggers      += YourPackageName
override LINK_JetUserObjects     += YourPackageName
override LINK_BTagAlgs           += YourPackageName
override LINK_BTagObjects        += YourPackageName
override LINK_BTagMods           += YourPackageName
```

and the following lines have to be at the end of the GNUmakefile:

```
ifneq ($(NEUROBAYES_DIR),)
         override CXXFLAGS += -DNEUROBAYES_DIR --backend -gstabs+
         override CPPFLAGS += -DNEUROBAYES_DIR
 include SoftRelTools/arch_spec_cern.mk
 include $(NEUROBAYES_DIR)/make_fragment/neurobayes.mk
 include SoftRelTools/arch_spec_f77.mk
endif
```

These lines have to be added to `YourPackageName/src/GNUmakefile` as well.
It also needed that the VxPrim module is linked to the executable. In order to achieve this, the following lines have to be included in the source build file for the executable:

```
#include "VertexMods/vxprim.hh"
...
aMod = new VxPrim();
add(aMod);
aMod->setEnabled(false);
```

The VxPrim module has to be enabled in the steering file and it has to precede in the path the module which calls the TrackProbNN class:

```
mod enable vxprim
# no talk-to for vxprim
...
path create fullPath ManagerSequence HepRootManager \
    vxprim \
    YourOtherModules
```

The executable is finally built in this way

```
gmake YourPackage.all
```

In order to run the executable, the expertise files have to be present in the directory from which the executable is launched.

H.3.1 Caveat

It might happen that the executable which calls the NNJetQTagger or the TrackProbNN class gives a segmentation violations in the tagger or in the track probability without any apparent reason. In this case one should make sure that BottomTaggers and the user package have been built consistently with the neurobayes setup. In the past crashes of these type where experienced and always been traced back to forgetting to set up the neural network package or to modify the file `YourPackage/src/GNUmakefile` and `YourPackage/test/GNUmakefile` as in Section H.3. If in doubt, a `gmake.clean`, the setup of neurobayes and the rebuilt of library and executable should help.

Bibliography

[1] S. Eidelman et al, *Phys. Lett.* B 592, 1 (2004)

[2] http://www-cdf.fnal.gov/physics/new/bottom/050310.bsmix-combined/

[3] http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/B/B20/B20.pdf

[4] CDF Collaboration, `hep-ex/0412057`

[5] T. Allmendiger et al., *"BSAURUS: A Package for Inclusive B Reconstruction in DELPHI"*, `hep-ex/0102001`

[6] Workshop on the CKM unitarity triangle (CKM2005), 2005
http://ckm2005.ucsd.edu/

[7] A.J. Buras, W. Slominski, H. Steger, *Nucl. Phys.* B 345, 369 (1984)

[8] T. Inami, C.S. Lim, *Prog. Theor. Phys.* 65, 297 (1981)

[9] A.J. Buras et al., *Nucl. Phys.* B 347, 491 (1990)

[10] A. Kronfeld, `hep-lat/0310063v1`

[11] T. Draper, *Nucl. Phys. Proc. Suppl.* 73, 43 (1999)

[12] CKMFitter, http://www.slac.stanford.edu/xorg/ckmfitter/ckm_results.html

[13] M. Battaglia et al., *CKM Matrix Workshop*, `hep-ph/0304132`

[14] C. Albajar et al., UA1 Collaboration, *Phys. Lett.* B 186, 247 (1987)
H. Albrecht et al., ARGUS Collaboration, *Phys. Lett.* B 192, 245 (1987)

[15] R. Barate et al., ALEPH Collaboration, *Eur. Phys. J.* C 4, 367 (1998)
A. Heister et al., ALEPH Collaboration, *Eur. Phys. J.* C 29, 143 (2003)
J. Abdallah et al., DELPHI Collaboration, *Eur. Phys. J.* C 28, 155 (2003)
P. Abreu et al., DELPHI Collaboration, *Eur. Phys. J.* C 16, 555 (2000)
W. Adam et al., DELPHI Collaboration, *Phys. Lett.* B 4141, 382 (1997)

DELPHI Collaboration, note 2002-073-CONF-607, contrib. 587 to Int. Conf. on High Energy Physics, Amsterdam 2002
G. Abbiendi et al, OPAL Collaboration, *Eur. Phys. J.* C 11, 587 (1999)

[16] K. Abe at el., SLD Collaboration, *Phys. Rev.* D 67, 012006 (2003)
K. Abe at el., SLD Collaboration, *Phys. Rev.* D 66, 032009 (2002)
SLD Collabboration, SLAC-PUB-8568, contrib. to 30^{th} Int. Conf. on High Energy Physics, Osaka 2000

[17] F. Abe et al., CDF Collaboration, *Phys. Rev. Lett.* 82, 3576 (1999)

[18] F. Abe et al., CDF Collaboration, *Phys. Rev. Lett.* 73, 225, 1994
S. Abachi et al., DØ Collaboration, *Phys. Rev.* D 50, 2966, 1994

[19] T. Affolder et al., CDF Collaboration, *Phys. Rev.* D 63, 032003 (2001)

[20] Courtesy of Ronald Moore, Fermilab Accelerator Division

[21] CDF Store Summary Web page, http://www-cdfonline.fnal.gov/opshelp/stores/

[22] CDF Collaboration, *"The CDF II Detector - Technical Design Report"*, FERMILAB-Pub-96/390-E

[23] CDF II Collaboration, FERMILAB-PUB-96/390-E (1996)

[24] A. Sill et al., *Nucl. Instr. Meth.* A 447, 1 (2000)

[25] T. Affolder et al., *Nucl. Instr. Meth.* A 42586, 249 (2004)

[26] D. Acosta et al., *Nucl. Instr. Meth.* A 518, 605 (2004)

[27] L. Balka at al, *Nucl. Instr. Meth.* A 267, 272 (1988)

[28] Y. Seiya at al, *Nucl. Instr. Meth.* A 480, 524 (2002)

[29] S. Bertolucci at al, *Nucl. Instr. Meth.* A 267, 301 (1988)

[30] G. Brandenburg, T. Liss, P. Schlabach et al., *"The CDF Run 1 Muon System Upgrade"*, CDF note 6362

[31] A. Bardi et al., *"The CDF-II Online Silicon Vertex Tracker"*, `hep-ph/0112141`

[32] C. Ciobau et al., *IEEE Trans. Nucl. Sci.*, 46, 933 (1999)

[33] C. Hays, P. Tamburello, A. Kotwal, P. Wittich, F. Snider, *"The COT Pattern Recognition Algorithm and Offline Code"*, CDF note 6992

[34] C. Hays, Y. Huang, A. Kotwal, H. Gerberich, S. Menzemer, K. Rinnert, C. Lecci, M. Herndon, R. Snider, *Nucl. Instr. and Meth.* A 538, 249 (2005)

[35] M. Feindt, S. Menzemer, K. Rinnert, *"TrackingKal - A Tracking and Alignment Software Package for the CDFII Silicon Detector"*, CDF note 5968

[36] M. Feindt, C. Lecci, S. Menzemer, K. Rinnert, *"Pre-Tracking PV z-Finder"*, CDF note 5988

[37] Y. Huang, C. Hays, A. Kotwal, *"Inside-Out Tracking"*, CDF note 6707

[38] H. Stadie, W. Wagner, H. Wenzel, T. Muller, *"Vxprim in Run II"*, CDF note 6040

[39] http://cdfkits.fnal.gov/CdfCode/source/JetUserObjects/JetUserObjects/ConeClusteringAlg.hh

[40] W. Yao et al., *"A Seed Vertexing b-Tag Algorithm for Top"*, CDF note 2716

[41] D. Acosta et al.,*"Introduction to Run II Jet Probability Heavy Flavor Tagger"*, CDF note 6315
A. Sukhanov et al.,*"Efficiency of Jet Probability Heavy Flavor Tagger"*, CDF note 6931

[42] http://www-cdf.fnal.gov/physics/new/bottom/050310.bsmix-semi/

[43] http://www-cdf.fnal.gov/physics/new/bottom/050310-bsmix-hadronic/

[44] P. Catastini et al., *"Same Side Kaon Tagging Studies"*, CDF note 7496

[45] V. Tiwari et al., *"Likelihood based Electron Tagging"*, CDF note 7121
G. Giurgiu et al., *"Muon B Flavor Tagging - A Likelihood Approach"*, CDF note 7043

[46] R.D. Field, R.P. Feynman, *Nucl. Phys.* B 136, 1 (1978)

[47] G. Bauer et al., *"B Flavor Tagging Using Opposite Side Jet Charge"*, CDF note 6951

[48] G. Bauer et al., *"Improved Jet Charge Tagger for summer conferences 2004"*, CDF note 7131

[49] H.G. Moser, A. Roussarie, *Nucl. Instr. Meth.* A 384, 491 (1997)

[50] I. Kravchenko et al., *"Measurement of B^0 Oscillations and Calibration of Flavor Tagging in Fully Reconstructed Decays"*, CDF note 7447

[51] Heavy Flavour Averaging Group,http://www.slac.stanford.edu/xorg/hfag/osc/index.html

[52] E. Norrbin and T. Sjöstrand, *"Production and hadronization of heavy quarks"*, hep-ph/0005110

[53] R. Field, *"The Sources of b-Quarks at the Tevatron and their Correlations"*, CDF note 5813

[54] T. Sjöstrand, *Comput. Phys. Commun.* 82 (1994) 74

[55] H.-C. Fang, A. Cerri, M.D. Shapiro, M. Tanaka, S. Uozumi, *"SemiLeptonicB/LeptonSvtSel - An offline filter module for lepton-SVT data"*, CDF note 6326

[56] M. Jones et al., *"Sample Composition of the ℓ+SVT Triggers"*, CDF note 6480

[57] M. Campanelli, E. Gerchtein, *"Calibration of the momentum scale for Kalman refitter using J/ψ events"*, CDF note 6905

[58] W. Bartel et al., JADE Collaboration, *Z. Phys.* C 33, 23 (1986)

[59] K. Pitts, *"Jet Charge Flavor Tagging in Fully Reconstructed B Decays"*, CDF note 4744
J.G. Heinrich et al., *"A Measurement of sin2β Using Multiple Flavor Tags"*, CDF note 4747

[60] I. Furić, *"Tuning of JQT with Mass Clustering"*, Semileptonic Subgroup Meeting, Aug. 3^{rd}, 2004

[61] M. Herndon, *"Datasets and integrated luminosity"*, BPAK Meeting, Sep. 24^{th}, 2004

[62] http://www-cdf.fnal.gov/upgrades/computing/projects/reconstruction/tracking/user-docs/l00.html

[63] B. Wicklund, *"Evaluation of Errors on ϵ, D, ϵD^2"*, CDF note 6716

[64] O. Long, *"A Proper Time Dependent Measurement of Δm_d Using Jet Charge and Soft Lepton Flavor Tagging"*, CDF note 4680

[65] C. Gross, Diploma thesis in preparation.

[66] http://grid.fzk.de

[67] J. D. Lewis, P. Avery, *"CLEOMC: The CDF interface to the CLEO Monte Carlo (QQ)"*, CDF note 2724

[68] http://www-cdf.fnal.gov/internal/physics/bottom/b-montecarlo/db/g020.txt

[69] W. Bell, J.P. Fernandez, L. Flores, F. Wuerthwein, R.J. Tesarek, *"User Guide For EvtGen @ CDF"*, CDF note 5618

[70] M. Feindt, *"A Neural Bayesian Estimator for Conditional Probability Densities"*, `physics/0402093`

[71] http://www.phi-t.de

[72] http://cdfkits.fnal.gov/CdfCode/source/BottomTaggers/

Acknowledgements

I would like to thank my supervisor, Prof. Dr. Michael Feindt, for accepting me in his group and for his guidance during the years of my graduate studies.
I would like to thank Prof. Dr. Günter Quast for the co-supervision of this thesis.
I would also like to thank Prof. Dr. Thomas Müller and Prof. Dr. Michael Feindt for letting me visit Fermilab in several occasions. Frau Edeltraud Haas and Herr Sven Fuchs have always been very helpful with travel organisation.
My work was mostly supported by a scholarship of the Graduiertenkolleg "Hochenergiephysik und Teilchenastrophysik".
I express my gratitude towards Dr. Gary Barker for his continuous support and Dr. Ilya Kravchenko for always being available to answer questions and to ask many. I am glad for having worked closely to Dr. Stephanie Menzemer, who always had friendly and positive encouragements. I am grateful to Dr. Kurt Rinnert for the "C++/CDF Software Helpdesk", which was precious in speeding up bug fixes and code development.
I thank the CDF B Group for the interesting discussions during the blessing procedure.
Several people contribute to my practical work. I would like to mention the Administrator Team of EKP and the people who helped me with the submission of CAF jobs in crucial moments.
I am grateful to the people that read this thesis and commented on it: Dr. Stephanie Menzemer, Dr. Markus Moch, Dr. Thomas Kuhr, Dr. Gary Barker, Dr. Kurt Rinnert, Jurriaan van Buren. Joachim Heuser, Claudine Groß, Yves Kemp and Dr. Stephanie Menzemer gave me valuable assistance in translating the summary to German.
Special thanks go to Thilo Wüst, Christian Sander and Yves Kemp, from whom I learned most of my German during the lunch breaks, and to the people at Fermilab who made my stays something to look forward to.
I would like to thank my family for never objecting to my choices.
I am most thankful to Jurriaan for his unconditional support and for being a constant source of inspiration.